Technical Document 3278
September 2013

Energy-Focused Fused Information System Integration

A NISE funded Capability Investment Project

Year 1 FY 2013 Report

Stacey Curtis
Adriane Wotawa-Bergen Wolfe

SSC Pacific
San Diego, CA 92152-5001

SSC Pacific
San Diego, California 92152-5001

J. J. Beel, CAPT, USN
Commanding Officer

C. A. Keeney
Executive Director

ADMINISTRATIVE INFORMATION

This report was prepared by the Research and Applied Sciences (Code 71) Department, SPAWAR Systems Center Pacific (SSC Pacific), San Diego, CA. This work was funded by the Naval Innovative Science and Engineering (NISE) Program at SSC Pacific as a Capability Investment project.

Released by
S. Curtis, Head
Environmental Applications Branch

Under authority of
M. Machniak, Head
Advanced Systems & Applied
Sciences Division

SB

CONTENTS

Figures

Tables

1. INTRODUCTION

Traditionally, Navy Command, Control, Communications, Computers and Intelligence Information Technology (C4I-IT) systems were developed for maximum computing performance and security and have not valued energy as a strategic resource. The goal of the Energy-Focused Fused Information System Integration (EFFIS-INT) project is to develop a capability that will enable Navy C4I-IT systems to optimize their design and operations in consideration of energy, computing performance, and cyber security. This capability investment includes a next-generation technical solution, business development, and workforce development. The technical product under development is a Navy-centric command and control (C2) energy management dashboard developed in Ozone Widget Framework to provide actionable information to end-users. Three main task areas will be used to develop the energy management approach: (1) data generation, (2) data analysis, and (3) dashboard widget development. For the experimental data generation, we established a virtual machine (VM)-based server computing environment to run benchmark loads. In the data analysis stage, the data generated via experimentation and modeling shall be analyzed using timeframe clustering and machine learning with an emphasis in fusing energy, performance, and security data for holistic system optimization. In the dashboard development, command and control ozone widgets are designed to display the data and the analysis results. We are using user-centered design to develop the dashboard design. In the first year (fiscal year 2013) of this effort, the technical thrust area was at the computing system level, but over the course of the project will be expanded to include energy storage, and cooling considerations to create a holistic approach for computing environments.

2. PROJECT OVERVIEW

The project overview section provides an overview of the high-level motivation and describes the project, including the technical problem, project objectives, and relevance to the Navy.

2.1 PROBLEM STATEMENT

IT systems are over-designed and, typically, operated significantly below capacity. Systems that are operating below capacity are inefficient, which wastes energy, capital investment, size, and weight. For example:

- Typically, only 30% of IT equipment is used.
- IT equipment is typically cooled at more than 10 degrees below the requirement.
- Often, less than 50% of expeditionary system generators are used.

2.2 PROJECT OBJECTIVES

The objective of this project is to develop sustainable science and technology (S&T) in command, control, communications, computers and intelligence (C4I) energy and energy management. Sustainability development speaks to development that can be sustained without compromising the future. The three pillars of sustainable development are economics, equity, and environment. In this project, three pillars of a capability investment were conceptualized to map to the pillars of a sustainable business development strategy (Figure 1).

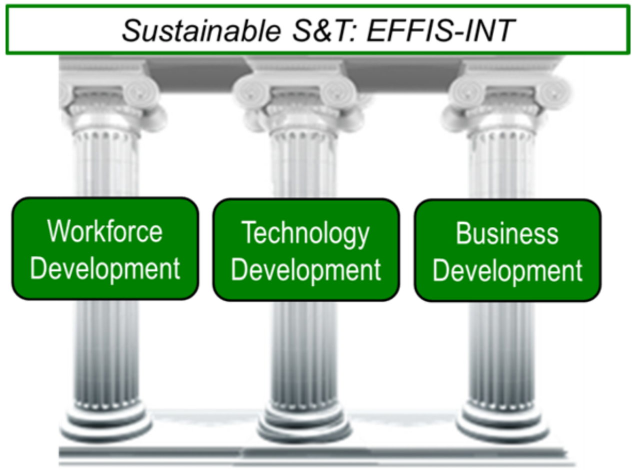


Figure 1. The three goals of a capability investment that correlate to the three tiers of sustainable development.

The specific goals are to conduct technical development of a command and control (C2) energy management system for Navy C4I/information technology (IT) systems. Three tasks were identified to achieve this development: (1) investigate computing, cooling, energy management and cyber physical security, (2) optimize the system to reduce energy usage while maintaining IT performance and security with the target of a 50% reduction in total system energy consumption; and (3) develop a C2 energy management approach for cross-platform Navy end-users. We have two business development goals: (1) grow a sustainable S&T business area related to this project, and (2) transition the EFFIS-INT project approach to C4I-IT systems. The goal of workforce development is to develop people who are technically knowledgeable in the field so that SSC Pacific can become a center of excellence in IT-energy management.

2.3 NAVY RELEVANCE

The U.S. Navy's 2020 goal is to reduce energy consumption by 50% and increase renewable energy by 50% at all its installations. The U.S. Marine Corps Expeditionary Energy Strategy seeks to reduce energy consumption by 50% for expeditionary systems by 2025. The Navy Expeditionary Energy Strategy wants to achieve 15% in fuel reduction by 2020. The Navy plans to leverage the Marine Corps as the leader in expeditionary energy reduction. In expeditionary systems, top energy loads are "life force and C4I."

This project is relevant to the Navy and mapped to specific thrust areas. The project is aligned with the SSC Pacific Technical Thrust Area, "R&D/S&T for Advanced Power and Energy Production and Efficient Use" and the SSC Pacific Special Emphasis areas of "Power and Energy Efficiency, Total Ownership Cost Reductions, Cyber Defense and Attack Automation, C2 of C2, Situational Awareness/Decision Awareness, Cloud Computing, Automated Multi-platform Control and Data Fusion." The project supports the Naval S&T Strategic Plan Technical Thrust Area, "Power and Energy, Total Ownership Cost, and Information Dominance." It also is aligned with the Deputy's Management Action Group (DMAG) guidance for a Comprehensive Defense Energy Policy and the SPAWAR Energy Reduction Strategy to reduce the energy footprint of the SPAWAR product line.

3. TECHNICAL RESEARCH & DEVELOPMENT

The technical approach is to (1) collect experimental and modeled data while enacting management policies, (2) conduct data analysis and develop system optimization policies, and (3) visualize and empower decisions using end-user-design dashboard development (Figure 2). Because this is a new technical area for SSC Pacific, before tasks 1–3 could begin, workforce and infrastructure development was required.

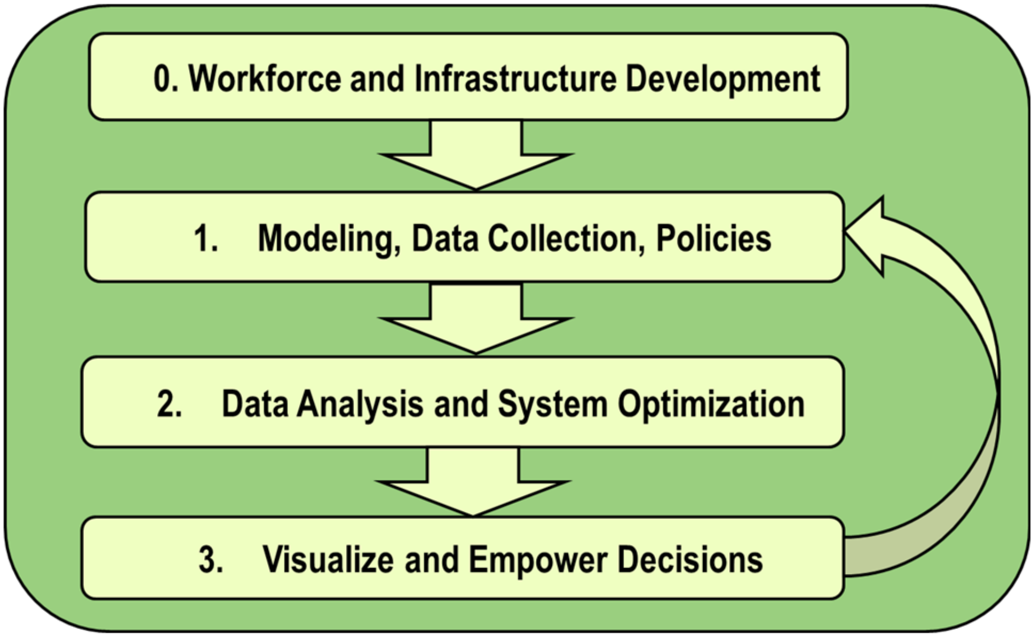


Figure 2. Technical approach flowchart.

3.1 TECHNICAL REVIEW

To conduct unique and meaningful research and development, a literature review of data center energy management was conducted with over 90 articles reviewed. The paper titled "Review of Scheduling-Based Data Center Energy Management" was submitted to *Elsevier Journal of Sustainable Computing* on September 16, 2013. A copy of the paper is in appendix A. Some key points were as follows:

Metrics

- Energy: total power, IT power, power utilization effectiveness, energy efficiency
- Computing: quality of service, service level agreement, utilization, VM
- Cyber: Poorly defined

Efficiency Approaches

- Turn on/off/sleep servers for IT efficiency
- Dynamic VM allocation for IT efficiency
- Dynamic voltage and frequency scaling for IT efficiency
- Scheduling IT workloads based on physical location for cooling efficiency
- Controlling the temperature and air flow for cooling efficiency

Reliability

- Tiers of service level
- Redundancy, capacity, backup

Power as Indicator

- Power spikes due to unusual compute utilization
- Experimental and modeling experimental approaches
- The usage of benchmark applications for simulated workloads

After the literature review, project members decided to take a holistic systems perspective for IT-energy optimization. For example, application of a data center system is defined in Figure 3.

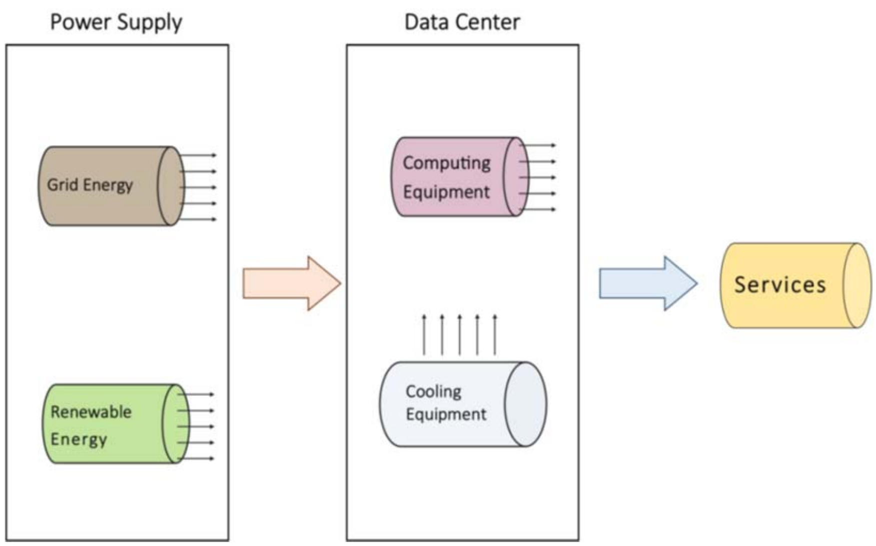


Figure 3. Data center systems perspective of energy where the input is power, the output is service, and the subsystems are energy, computing, and cooling.

The goal of this project is to develop holistic energy management for Navy IT systems. The specific management goals associated with the system should describe energy management. For the EFFIS-INT project, our two primary goals are energy efficiency and energy security. The data center can consider increasing energy efficiency through IT efficiency, cooling efficiency, or renewable insertion, where each of these approaches reduces dependence on the grid and associated financial and greenhouse gas emission costs (Figure 4). From an energy security perspective, security threats can come from a reduction in grid availability, a reduction in renewable availability, and an increase in the power requirement of the IT loads (for example, due to CPU-based denial of service attack) (Figure 5).

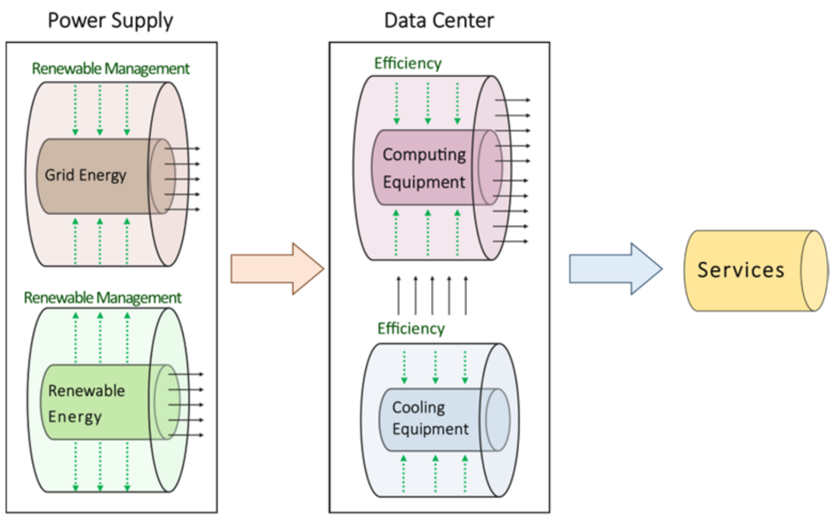


Figure 4. Energy efficiency of systems for data centers.

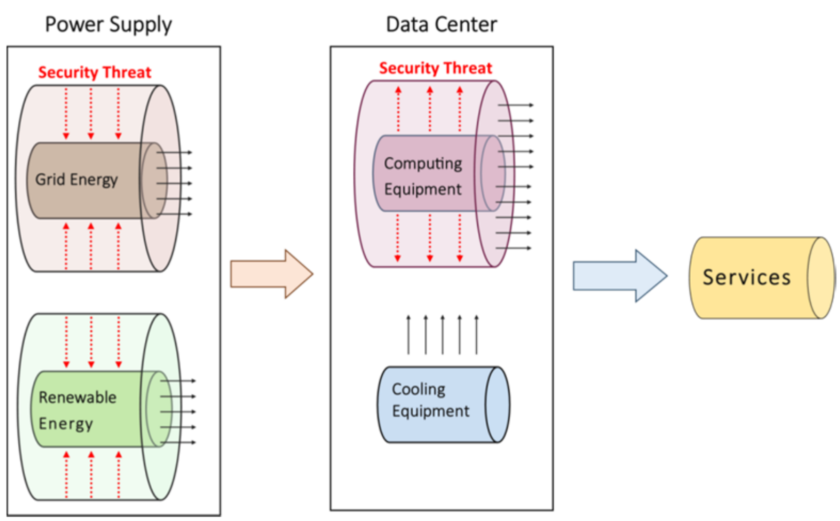


Figure 5. Energy security of systems for data centers.

3.2 DATA GENERATION AND COLLECTION

The goal of data generation and collection is to generate data to analyze the relationship between energy, performance, and security to optimize and manage IT systems more effectively. In FY 2013, we focused on establishing data generation and collection techniques and then practicing key management techniques to develop an energy-management toolkit. The idea was that these techniques can be used optimally to achieve IT efficiency. From the literature review, four primary IT efficiency management techniques were identified (Table 1). In FY 2013, we investigated and conducted VM dynamic scheduling and dynamic voltage and frequency scaling. Project personnel are developing server mode scheduling and power capping capabilities in another current project (DC Smart-E) and these capabilities will be leveraged in the future so that management policies can consider including all four techniques.

Table 1. Compute efficiency management techniques.

Compute Efficiency Management Techniques	
Technique	**Variable**
Server Mode Scheduling	On/off/sleep
VM dynamic scheduling	VM allocation
Dynamic voltage and frequency scaling (DVFS)	Frequency
Power Capping	Device power

The laboratory setup and VM loading experimentation conducted in FY 2013 shall be presented in an IEEE conference proceeding and is included in Appendix B. Key development included the following:

- The setup of the Energy Innovation Laboratory that houses the Computing Energy Testbed (CET) for experimental testing.
- The test bed includes smart metering leveraged from the Data Center Smart Metering Evaluation project; however, those data collection methods were less effective for the sensitive, accurate, granular data collection required to make performance/energy comparisons, so an alternative was found using a power quality analyzer (Hioki) and VMware.
- MATLAB®-based modeling of expected performance based on VM loading utilization and dynamic voltage and frequency scaling (DVFS)
- Experimental results for VM loading utilization
- MATLAB®-based data analysis, identification of key parameters, and comparison of experimental and model performance.

Not included is the initial DVFS experimentation where the impact of utilization and frequency are considered from an energy usage perspective. The reason is that the experiments have not had power performance as expected based on the driving equation related power to frequency. Continued investigation and analysis is required before results can be shared externally.

In addition, a high school summer intern from the Science and Engineering Apprenticeship Program (SEAP) developed an independent student research project to be submitted to the Siemens Science and Engineering competition. The student report is included in Appendix C. Key work included the following:

- Selecting a data center workload model
- Designing a hybrid photovoltaic-battery system
- Designing energy management techniques for specific grid pricing models
- The calculation of energy cost savings, capital investment, and system return on investment (ROI)

Future work shall focus on continuing to experiment with the "toolkit" of energy management techniques, including the refinement of DVFS and dynamic load shifting of VMs between physical servers. Data will be collected while testing this toolkit to understand the impact of using techniques on energy, performance, and security. The goal is to define policies for managing these techniques without compromising service-level requirements.

3.3 MACHINE LEARNING

In addition to standard data analytic approaches, both supervised and unsupervised machines were applied to real-world and simulated power usage data to automatically classify usage patterns and detect variations to generate an appropriate response to a particular situation. Initial results showed that traditional clustering approaches were less successful than artificial neural networks trained through evolutionary computation or deep learning. Task 2 allows for greater understanding of the energy challenges faced by the U.S. Navy and the Department of Defense (DoD). By incorporating machine learning into the analytics, such decisions can be optimized and routine tasks that maximize energy efficiency can be automated. Additionally, autonomous power monitoring can be used to detect anomalous behavior that may be indicative of cyber threats. Machine learning approaches have been applied to time-series prediction problems for a long time (Weigend, 1994). Such problems involve predicting the possibility of future events based on a history of related events. An emerging trend for analyzing these problems with complex multimodal data (such as that seen in power monitoring) is deep-learning neural networks, especially when used as auto-encoders (Hinton et al., 2011).

For Task 2, data analytics, several machine-learning techniques were surveyed and evaluated to identify anonymous behaviors that do not follow predicted trends. Initial efforts focused on unsupervised clustering approaches: mixture of Gaussian and k-Nearest Neighbors. These initial experiments in anomaly detection showed issues with these traditional techniques. Specifically, many parameters in the power data and many of these parameters had narrow ranges of values, resulting in sparse data and over-classification of anomalies. An alternative approach to the problem is to employ machine learning to model and predict usage trends. To this end, the NeuroEvolution of Augmenting Topologies (NEAT) algorithm was used to create artificial neural networks (ANNs) that could predict resource usage. Due to limited available data, a multimodal simulation that mimicked real-life usage patterns of computational resources was created to assist in training. NEAT created neural networks that could accurately predict future usage based only on past power consumption. However, the neural networks produced by NEAT displayed complex internal structures, implying that deep-learning neural networks may be a more appropriate choice. Initial deep-learning tests support this conclusion. The current focus of Task 2 is the implementation, refinement, and testing of deep learning neural networks for predicting future resource usage and to determine if current usage is anomalous/requires attention. Task 2's future work will continue to refine the predictive analytics through deep learning as more data becomes available. Additionally, it will expand its focus to exploit relationships between power usage and cyber-attacks to provide prediction and altering of such threats.

This neural network can predict future processor load based on historical power consumption with 99.2% accuracy. The network was trained with the NEAT algorithm on raw power data from a server running increasing numbers of virtual machines. The complex internal structure implies that deep-learning architectures may be a more appropriate approach. Figure 6 shows the normalized power consumption (black) over time and the neural network's predicted processor usage (red, normalized out of 8) at the same time. Note that the normalized processor usage and power usage should not necessarily be the same value.

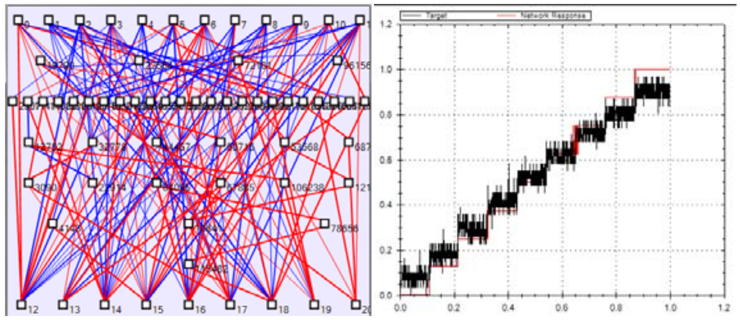


Figure 6. Evolved neural network.

3.4 VISUALIZE INFORMATION AND EMPOWER END-USERS

A management dashboard is critical to communicating information to end-users and enabling management or control functions. In FY 2013, the project focused on developing dashboard concepts for data center management, including end-user scenario definition, end-user interviews, conceptual dashboard design, and C2 implementation documentation. In addition, project personnel have thought about initially focusing on expeditionary end-users and how EFFIS-INT could be tailored to their needs in FY 2014.

3.4.1 Data Center End-User Scenario

Data Center Joe is a data center manager and must report on the current operating cost of energy to his line management. He is also required to make cost-saving suggestions. Joe decides to integrate the EFFIS-INT C2 Optimization system into his data center. This optimization system will allow him to collect data on the physical environment, computing performance, and power and energy information. Joe could then use the tool to establish a baseline on the current system and determine current costs. To help with cost saving, Joe could use the optimization aspect of the tool. This optimization would have several effects. The first would be that it would analyze data to figure out which systems could be automatically improved. An example would be shifting the load onto more servers located in a cooler part of the lab. In addition, the system would advise on actions the user could take to further increase optimization. Joe can now compare the savings on the optimized system to the baseline. Using these numbers, he can support his decisions to continue making improvements. Joe is also responsible for maintaining operations 24/7; the optimization tool supports this by conservatively applying policies to avoid violations of service-level agreement. When an imbalance exists between power availability and power requirements, the system prioritizes workloads based on operational priority and risk/optimization trends developed through the machine-learning function of the software.

3.4.2 Data Center End-User Interviews

Throughout FY 2013, we have conducted nine interviews of potential end-users. The demographics for the participants is as follows:

Ages: 25-65

Education: 9 had a higher level degree

Prior Military Experience: 3

Roles: System administrators, data center operators, configuration management/transition support for data centers, facility operations managers, adviser for electrical systems, and a research investigator

The topics discussed were general definitions of efficiency and optimization, energy efficiency, cyber security, details about the tasks they currently do, and information about the tools supporting those tasks. Broad topics were used to collect information on these subjects and then use that information to guide the direction of our solution.

Here is a compiled list of the key findings:

- Trust - system must prove to users that the data is accurate and the system is reliable
- Transparency - calculations and process must be visible or accessible to users to build trust
- Energy Efficiency Mentality - 7/9 are only concerned with energy efficiency because they are ordered to be concerned about it (usually to save money)
- Automation - automating many tasks would be very useful (e.g., load sharing)
    - Capability to set times for it to not occur
    - Ability to override
- Alerts - provide feedback when different events occur (e.g., generator getting over loaded, cyber-attack)
    - Allow user customization—excessive alerts would get annoying
- Permission structure - control is needed over who can do what, but need a way to override certain things as well
- Integration with existing management systems used
    - Solar winds, Symantec Net Backup, inMon, Zenoss, Remedy
- Pros of systems used (summary of all pros and not applicable to all systems)
    - Easy to learn
    - Extensible
    - Filtering
    - Responsive
- Cons of systems used (summary of all cons and not applicable to all systems)
    - No easy integration
    - Hard to set up
    - Clunky navigation
    - Not configurable
    - Too much drill-down required
    - Not intuitive, learn by trial/error
- Data gathering: must walk around the space and use disconnected tools to gather certain information (e.g., airflow or spot temperature)

- Equipment management: understand what equipment will be used at different locations so resources can be better allocated
- Incentive: currently not much incentive for trying to be energy efficient
- Return on Investment (ROI): need to show ROI in ways leadership can understand and appreciate
- Modeling: real-time modeling capabilities to try different settings and configurations would be extremely useful
- Remote Access: need capabilities to manage in case of evacuation or other physical access issues
- Adaptability: the system should work with different technologies and manufacturers
- Networks: need to manage both classified and unclassified systems
- Approval: system should move items up the approval chain and be traceable
- Physical layouts: different maps, schematics, and blueprints should be available
- Cyber Security—isn't a major concern to most users, but know it is important
  - Cyber secure systems are essential
  - Physical systems can be affected
- Data Types: the most mentioned data types were environmental conditions, power usage, and computing information; Cyber data was not brought up unless for an alert of an attack.

3.4.3 Conceptual Dashboard Design

A conceptual dashboard was designed in wire-diagram form, as seen in Figure 7 and 8. Individual demo Widgets are documented in Appendix D. An overview of the dashboard software requirements for C2 type systems is in Appendix E.

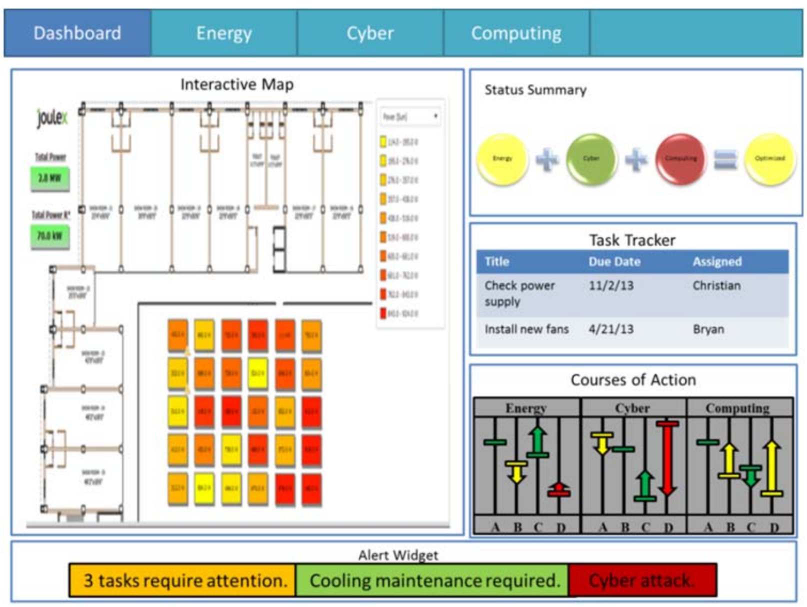


Figure 7. Dashboard diagram A.

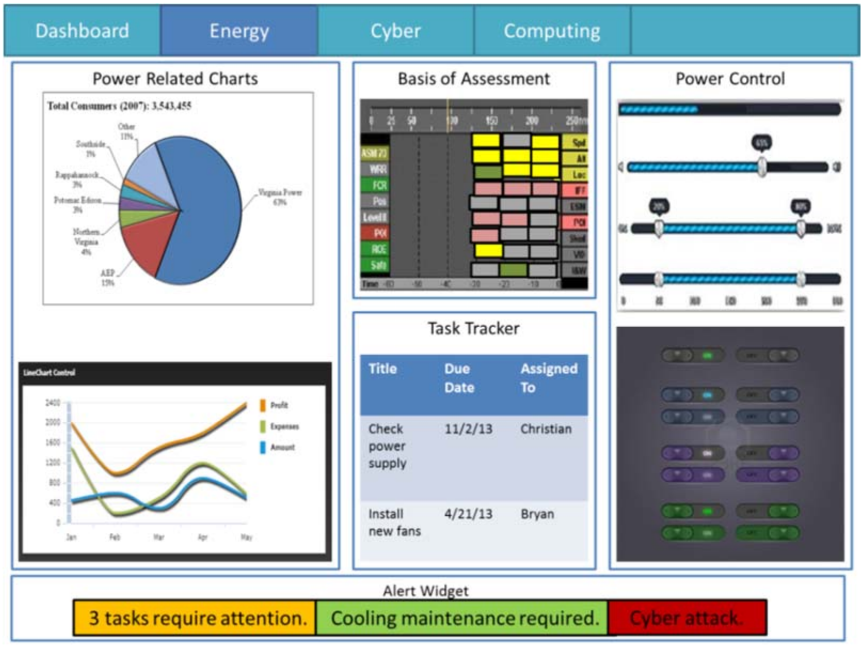


Figure 8. Dashboard diagram B.

3.4.4 Expeditionary End-User Scenario

An alternate scenario was developed for expeditionary systems where the environment, end-user, problem, technology, and solution are defined. When key information was not available, notes were made to highlight the need to identify the information.

3.4.4.1 Environment

The operating environment is a forward operating base (FOB) that is at the scale of 10 tents, 10 generators, with a fixed position for at least 12 months. The FOB has a command and control (C2) room and a number of loads that require reliable power: hosts one to five big screen projectors, five to ten tough book computing work stations, one to five server racks, and a cooling system. The C2 space is divided into sub-rooms or sections for work groups to be isolated for communication security. The C2 room is either a tent with raised floor or an armored trailer called a silver bullet. The source of the power is a JP8 fuel diesel generator that provides power for the cooling and IT loads and is rated for 10- to 50-kW peak power. The racks have rack-mountable uninterruptable power supplies (UPS) to provide short- term energy reliability for the servers. The IT servers and tough books are classified and are operated on the SIPERNET.

Questions for refinement:

- Is the FOB scale correct?
- Is the C2 room equipment scale correct?
- Is the generator size scale correct?
- How much power to the devices typically use/ are rated for?
- Are their generally rack mounted UPS?
- Are all the systems SIPERNET?
- How centralized are the power systems?

- How are cables run from building to building? (e.g., if you have a personnel tent away from the C2 tent and it has its own generator, would cables networking all these areas together create extra problems?

3.4.4.2 End-User

Frank is the equivalent of an Army 35T technician at the forward operating base. He is responsible for a wide range of technical support, including supporting the IT and power systems. Frank works in a small group led by Warrant Officer Joe. The warrant officer provides information to the supply sergeant for fuel supply requirements.

Questions for refinement:

- What are the equivalent positions for the U.S. Marine Corps or U.S. Navy?
- How does the decision hierarchy work? Are these the correct positions?
- What are Frank and Joe's ranks so we can call them by their titles?
- What are some of Joe and Frank's key tasks?
- How important is resource management, energy conservation, etc., to these people?

3.4.4.3 Problem

Fuel supply is a significant challenge for a forward operating base; in Afghanistan, the fully burdened cost of fuel was $300 and fuel convoys were one of the largest security risks to our expeditionary forces. Because fuel convoys are both expensive and risk lives, the U.S. needs to reduce the amount of fuel convoys and fuel consumed by forward operating bases. Currently, Frank does not have the tools or expertise to optimize the efficiency of fuel consumption and manages a setup that is over-designed. Generators and servers operate inefficiently, with utilization estimated at from 10 to 30% of the system's capacity. Frank and Joe do not feel comfortable reducing the overshoot because they are concerned that they will require the additional capacity if a critical event occurs. Running systems at this low level of utilization can lead to losses of 50% on both the supply and demand side, and combined, can result in total systems with only 25% efficiency. However, Frank and Joe are unaware of how inefficient their system is or that it is requiring them to request refueling four times more often than would be necessary if the system was 100% efficient in its energy management. With no information on how the systems are operating, Joe does not know when to tell the supply sergeant to request more fuel or how much fuel to request, so he orders it more often and at larger quantities than necessary to ensure the reliability of the systems.

In addition, Frank has been using multiple disconnected systems to try to manage these resources and is regularly required to change the setup to avoid system outages caused by spikes in utilization. Frank does not have any tools to manage these spikes on the load or supply side. This leads him to request additional generators to meet the requirements during the spikes in usage.

Questions for refinement:

- What is the average utilization of the systems?
- How inefficient are the systems at this utilization?
- How often are re-supplies ordered?
- How do they determine when fuel will be supplied?

3.4.4.4 EFFIS-INT

To help with this problem, Joe requests the new information assurance (IA) certified and validated EFFIS-INT C2 Optimization System for Frank to implement. EFFIS-INT is energy management software that will be installed on one of the existing servers or a computer and collect energy data, analyze it, and provide Frank and Joe with a simple C2 dashboard. The EFFIS-INT system can collect utilization and power data in two ways, it can pull data via a network from devices or device meters, or if a type of information cannot be automatically collected via the network, Frank can manually enter the general device specifications or usage information he collects. This will allow for the management of legacy generators, etc. The EFFIS-INT system can analyze the efficiency and reliability of the IT energy load, cooling energy load, Universal Power System and other energy storage, the generator supply, and management of energy of the total system. If the network goes down or a device fails to report, the system will go in off-line mode and estimate status based on previous conditions, run-time, and end-user input. The C2 Dashboard will communicate critical information to Frank and allow special control functions based on user-level permissions.

3.4.4.5 Using EFFIS-INT

Right away, Frank will see how the power, cooling, and computing systems are currently used. The system will help him estimate how long their current resources will support their efforts. To make things better, Frank decides to use the optimization aspects of the system. EFFIS-INT will analyze the current data and optimization will occur in a couple of phases. First, the system will handle any aspect (i.e., server load) that can automatically be optimized. Next, Frank will be presented with a list of actions that could optimize the system. Some of these listed items will be handled by the system but require elevated permission, and others will require a physical action. For instance, the system could tell Frank that he is currently running two generators at a sub-optimal capacity. Frank would then shut down one of the generators and switch the items over to the other so he could reduce his fuel cost. The system will also provide supply logistics recommendations such as when to order more fuel and how much to order. It will also remind Frank to change the air filter and make other minor changes for efficiency. With a more optimized system, the resources Frank has to work with will last longer while still providing the level of support needed. The reduction in fuel would reduce the logistical cost of supporting forward-deployed systems.

4. WORKFORCE DEVELOPMENT

The workforce development aspect of this project was particularly intensive in FY13 because this is a new area for SSC Pacific, and none of the team members had previous expertise in the exact aspects of this project. Each team member brought unique experience, ranging from cloud computing, cyber security, machine learning, command and control, and energy systems. The team worked collaboratively, sharing their respective field expertise to develop this project. The FY13 team is shown in Figure 9.

The team started with six SSC Pacific government employees and grew with an increase in SSC Pacific participants as well as interns and touring professionals (Figure 10).

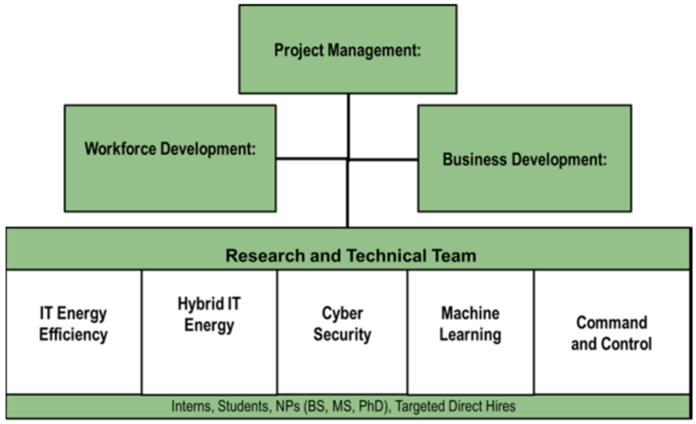


Figure 9. Organizational chart describing the team roles.

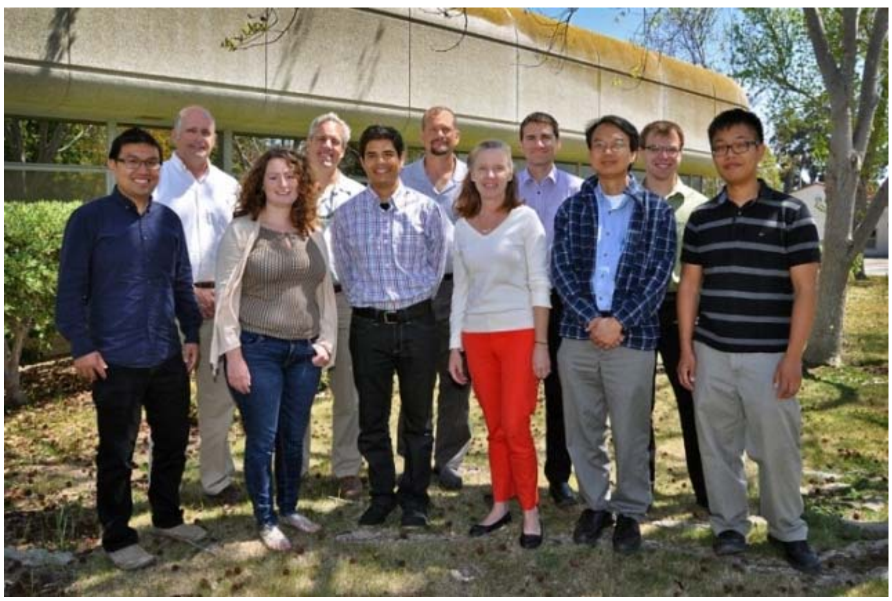


Figure 10. EEIS-INT Team.

In addition to adding new participant and training them in IT-energy management, we recognized that our team did not have enough individuals taking a leadership role in the project and adjacent projects that were part of the business development. For the projects started in FY 2013, five instead of three people will be taking a leadership role in FY 2014 (Figure 11). In addition, three more people were involved in leadership roles in proposed efforts to start in FY 2014. This shows not only a growth in team size, but reflects mentoring, development, and the interest of participants in developing their expertise in this field.

The project's goal is to develop a sustainable workforce during the next 5 years (Figure 12), which depends on the project's rapid growth in FY 2013 and plans for strategic sustained growth in leadership and subject-matter expertise. The size of the project team is expected to decrease as the internal funding tails off, but people will continue to grow in the field by working on related externally funded efforts.

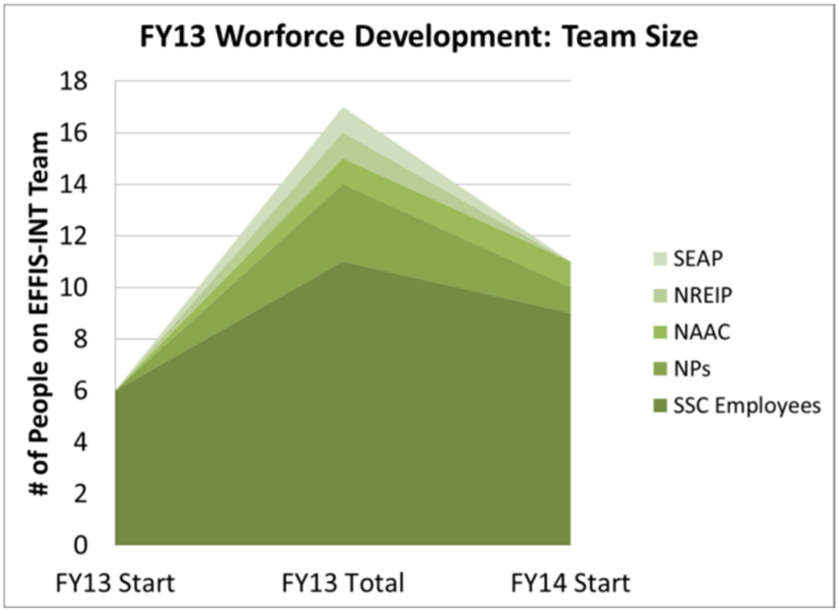


Figure 11**.** FY 2013 workforce team size.

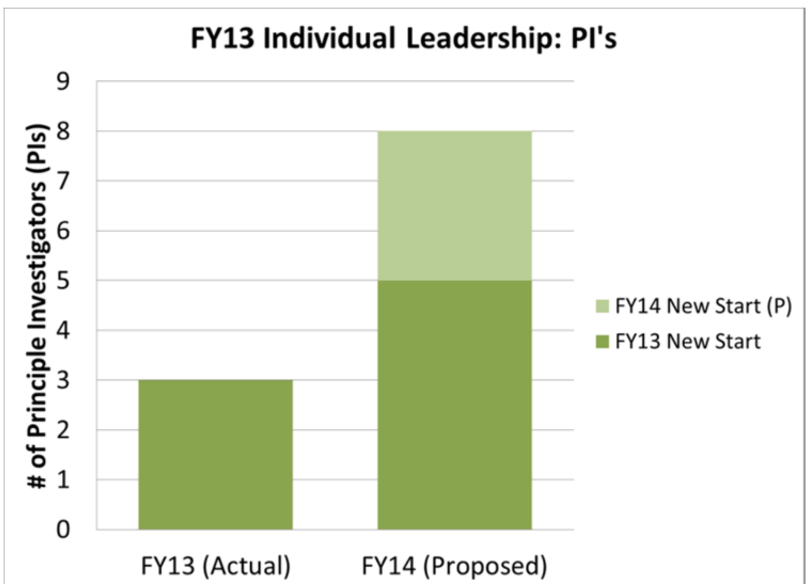


Figure 12. Growth of individuals taking leadership roles in IT-energy.

The project's goal is to develop a sustainable workforce during the next 5 years (Figure 13), which depends on the project's rapid growth in FY 2013 and plans for strategic sustained growth in leadership and subject-matter expertise. The size of the project team is expected to decrease as the internal funding tails off, but people will continue to grow in the field by working on related externally funded efforts.

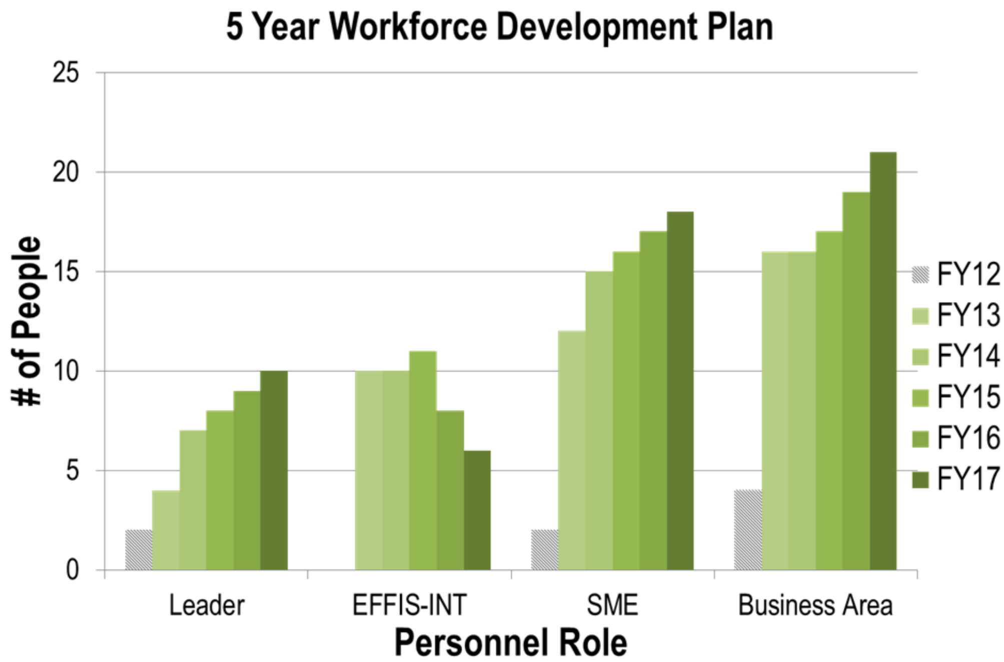


Figure 13. 5-year personnel growth plan.

5. BUSINESS DEVELOPMENT

Business development involves two areas: (1) broad area business development, and (2) project business development.

5.1 BROAD AREA BUSINESS DEVELOPMENT

Through the process of proposing the EFFIS-INT project and through relationships with the Data Center Consolidation Energy Team, adjacent projects were proposed in FY 2012 to complement the EFFIS-INT project (Figure 14). This resulted in external funding starting in FY 2013 from funding sources of previously unavailable funds for the project, and therefore, considered new business. Funding for many of those projects was front loaded because of the high cost of energy technology acquisitions. Those projects are scheduled to continue at lower levels of funding in FY 2014. In addition, additional funding was proposed for new projects for FY 2014. All of the projects of technology adjacencies with the EFFIS-INT project and the findings conducted in those projects will be used to inform others involved in the development of the EFFIS-INT project.

The “adjacencies” previously mentioned were formally designated as business development thrust areas. These thrust areas are critical to the success of the IT-energy area and the EFFIS-INT project. Team members participating in these projects further develop their technical expertise and benefit the EFFIS-INT project. The thrust areas are IT management, cooling management, energy management, and cyber–physical management (Table 2).

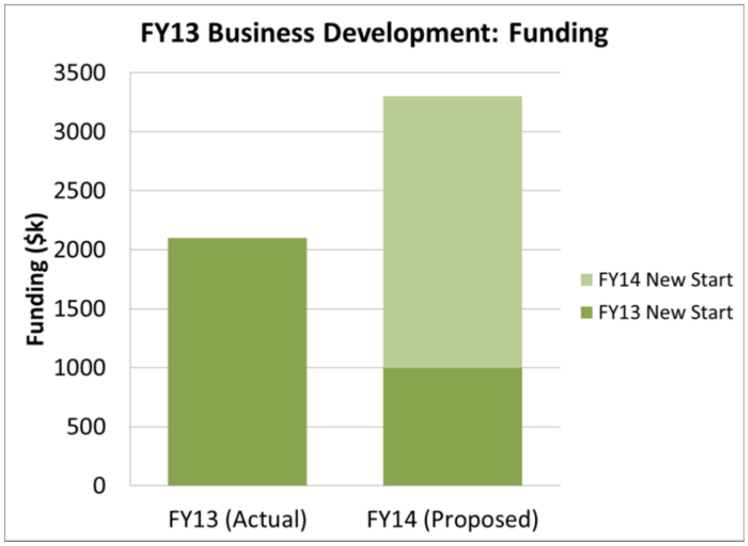


Figure 14. Funding for FY 2013 (actual) and FY 2014 (proposed).

Table 2. FY 2013 business development thrust areas.

FY13 Business Development: Thrust Areas				
	IT Management	Cooling Management	Energy Management	Cyber-Physical Management
FY13 New Start	1	2	3	2
FY14 New Start (P)	2	1	4	0

Each specific project is listed in Tables 3 and 4. These tables highlight that while complementary, each is distinctly different in its goal, tasking, and funding.

Table 3. FY 2013 and FY2014 Energy C4I /IT funding and sponsors.

Project	EFFIS-INT	Data Center Smart Metering	SCADA Cyber Security	Virtual Net-Zero Energy Planning	Data Center Cooling Efficiency
IT Management	FY13: Server Virtualization & Dynamic Voltage Frequency Scaling	Server On/Off/Sleep			
Cooling Management	FY14: Server temperature/VM load management	Computer Room Air Conditioner (CRAC) , room temperature			Cooling computational fluid dynamics air flow modeling & Physical containment
Energy Management	FY14: Server & battery/UPS management	Server, uninterruptable power supply, Rack power, DCIM management software		Power distribution modeling, renewable source and storage planning	CRAC Power, Rack power for waste heat predictions
Cyber-Physical Management	FY15: Denial of Service	Network Security Modeling	Posture of IA Offerings Evaluation		
Sponsor/ Funding FY13	NISE 6.1-6.3	ESTEP 6.3	ESTEP 6.3	ESTEP 6.3	Directed Energy 6.4
Project Type	Research & Development	Evaluation			Demonstration/ Validation

Table 4: FY 2014 proposed energy innovation new starts.

Project	Broad Agency Announcement with SSC technical management	Energy Efficient Cloud Computing Architectures	Advanced Energy Storage	RFID Smart Metering	Bosch Storage and Energy Management System
IT Management	DC powered servers	Cloud computing, energy of computing			
Cooling Management	Liquid cooling of servers				
Energy Management	Energy systems with hybrid and renewable management		Energy system optimization	Sub-building level metering using RFID meters	Power distribution modeling, renewable source and storage planning
Cyber-Physical Management					
Sponsor/ Funding	ESTEP 6.3	ESTEP 6.3	ESTEP 6.3	ESTCP 6.4	ESTCP 6.4
Project Type	Evaluation			Demonstration/ Validation	

The EFFIS-INT project focused on efficiency research in FY13 while having the goal to look at both security and efficiency. The below description shows how the research conducted in the Cyber Supervisory Control and Data Acquisition (SCADA) project (that was developed in FY13) can be leveraged in the future. This demonstrates a leapfrog-type business model where we build and feed off of adjacent efforts, enabling us to develop a capability during the next 5 years that will meet our aggressive project goals (Figure 15).

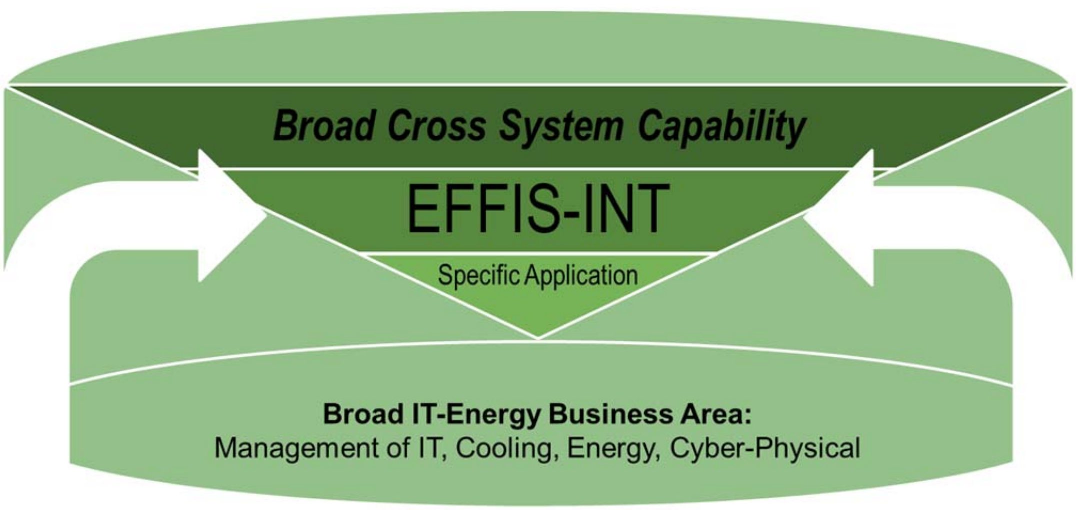


Figure 15. Strategic business development.

5.1.1 Cyber SCADA

Industrial Control Systems (ICS) are computer systems consisting of hardware, software, and communication components that enable monitoring and control of various infrastructures and systems. Supervisory Control and Data Acquisition (SCADA) networks are one type of ICS, designed to monitor and control critical infrastructures in electrical, water, oil, gas, and data industries. The nature of SCADA networks makes them extremely vulnerable to security threats, and more importantly, cyber-attacks. Today, no requirements can be applied across systems, infrastructures, and commands; this is a major problem for the DoD as existing SCADA networks are virtually unknown and virtually pieced together. Furthermore, no established metrics exist for measuring the security of such networks. As a result, SCADA networks today are not only vulnerable, but information about them is extremely limited to operators and decision makers.

The main objective of C-SEC (Cyber-SCADA Evaluation Capability) is to support a much more secure posture of SCADA networks by enabling the proper and repeatable evaluating of technologies that offer support for securing them. Furthermore, this secure posture will be accomplished through understanding Cyber SCADA needs, evaluating existing solutions for monitoring and protecting against cyber threats, and developing a new Cyber SCADA strategy. While we will research various types of systems, we will focus on energy systems. This project is expected to accomplish the following sub-objectives:

- Develop testable and measurable cyber security requirements for SCADA networks
- Devise a set of metrics for measuring the level of support existing (and future) technologies offer for securing SCADA networks
- Develop a repeatable and streamlined process for evaluating support offered by security SCADA technologies
- Develop a laboratory environment to baseline current cyber SCADA capabilities
- Evaluate various technology options for cyber SCADA monitoring and protection and provide a score, which represents their level of support
- Determine relationships, tradeoffs, and challenges of security and energy for SCADA
- Report on the state of the art of Cyber SCADA networks and outline gaps and recommendations for improving their cyber security posture

The team is researching this area and developing requirements as well as metrics for measuring the security level of SCADA networks. The establishment of a Cyber SCADA Laboratory, which will house a controlled small-scale SCADA network, is accomplishing this task. Using this network, the team will learn about vulnerabilities, produce data that alerts users about these vulnerabilities, and ultimately understand how to secure these types of networks.

The requirements and metrics developed as part of this effort will feed directly into the EFFIS-INT dashboard configuration and provide relevant data points for correlation. With information extracted from SCADA networks, EFFIS-INT will leverage all of the other information collected to aid in detecting possible cyber-attacks that would go completely undetected if not for something like EFFIS-INT. For example, zero day attacks (attacks that detection systems have never seen before) are an everyday concern for DoD networks; the problem is that the traditional security strategy of constantly reacting to cyber-attacks has put DoD in a position where it is hard to predict attacks in the first place. Information like CPU energy usage could be used to detect minimal (and most likely abnormal) changes in computing infrastructures that otherwise might not be caught using traditional defensive mechanisms.

5.2 EFFIS-INT BUSINESS DEVELOPMENT

While broad development was the focus of FY13, consideration and development was given to the growth of the EFFIS-INT project and its strategic growth.

In the Report of the Afghanistan Marine Energy Assessment (2011), the cost of energy was \$6/gallon at “operational” bases, which consumed 150,000 gallons/day, accounting for \$300 million a year while the cost of energy was more than \$300,000 a gallon at “tactical” bases, which consumed

30,000 gallons a day, and accounted for $320 million a year. In addition, the document highlighted that one U.S. Marine Corps casualty occurred on average for every 50 fuel convoys to “tactical” bases. In this report, the authors noted that two largest energy sinks are life support and C4I systems. For example, the Marine Air Ground Tactical Force experienced a 300% increase in IT/computers in 10 years. The Initial Capability Gap Assessment for U.S. Marine Corps Expeditionary Energy (31 Jan. 2011) highlighted the gap of plan for reduction in energy demand, plan to produce C4I energy onsite, and the need for energy-efficient climate control. These gaps were mapped to warfighter functions, including command and control and intelligence. Based on the following drivers, it was determined that the EFFIS-INT project will focus on an expeditionary-oriented solution for FY 2014.

6. CONCLUSION

In conclusion, we have created the beginning of a sustainable S&T business area, including a meaningful technical area, people with a passion for developing the area, and have acquired related funding. A summary of the FY 2013 accomplishments highlighted in this document are as follows:

Technical Highlights

- Literature Review Nov 2012- Feb 2013 (Appendix A)
- Energy Innovation Lab equipment setup Apr. 2013 (Appendix B)
- VM loading experimentation and analysis Jun-Jul 2013 (Appendix B)
- DVFS experimentation Aug-Sept 2013 (Section 2.2)
- Hybrid energy management modeling June-Aug 2013 (Appendix C)
- End-user interviews Dec 2013- Mar 2013 (Section 2.4)
- Scenario development Apr-Jun 2013 (Section 2.4)
- Dashboard conceptual design and documentation May-Aug (Appendix D, E) Publications
- Publications
  - A. Wolfe, et al, "Review of Schedule-based Data Center Energy Management," *Elsevier Journal of Sustainable Computing*, Submitted September 16, 2013.
  - B. Kwai, et al., "Hybrid photovoltaic energy management for data center application: A look toward cost competitive solutions." Siemens S&T High School Paper Competition, drafted for submission, September 20, 2013.
  - A. Wolfe, et al, "Computing Energy Testbed: A Model for the Evaluation of Energy Management Techniques for Data Centers," IEEE International Conference Cluster, Cloud, and Grid Computing," drafted for submission due November 11, 2013.
  - SSC Pacific, article series, "Energy Innovation Lab," drafted for submission, September 17, 2013.
- Patents
  - Adriane Wotawa-Bergen and Doug Lange, "Computing Systems Smart Metering Planning Widget," U.S. Patent Disclosure, April 15, 2013.
  - Adriane Wotawa-Bergen, Bryan. Croft, Christian Szatkowski, Jose Romero-Mariona, and Doug Lang, "Energy Efficiency and Energy Security Optimization Dashboard for Computing Systems Design," U.S. Patent Disclosure, March 25, 2013.
- Presentations
  - PMW 790 APM, PEO C4I Tech Match, OPNAV N45 Navy Energy Coordination Office (NECO), Navy International Programs Office (NIPO), ONR and ONR Global, NAVFAC Smart Power Partnership Initiative (SPPI), Task Force Energy (TFE)
- Documentation
  - FY 2013 Annual Report
  - Software Document

APPENDIX A: Review Paper

Sept. 16 2013 Submitted to:
Elsevier Journal of Sustainable Computing

Title:

Review of scheduling-based data center energy management

Authors:
Adriane Q. Wolfe
Space and Naval Warfare Systems Center Pacific
53560 Hull Street, San Diego CA, 92152-5001

James J. Jen
Space and Naval Warfare Systems Center Pacific
53560 Hull Street, San Diego CA, 92152-5001

Daniel Lyons
Department of Mathematics and Statistics
San Diego State University
5500 Campanile Dr. San Diego, CA 92182-7720

Chris S. Chen
Space and Naval Warfare Systems Center Pacific
53560 Hull Street, San Diego CA, 92152-5001

Vincent V. Dinh
Space and Naval Warfare Systems Center Pacific
53560 Hull Street, San Diego CA, 92152-5001

Jose Romero-Mariona
Space and Naval Warfare Systems Center Pacific
53560 Hull Street, San Diego CA, 92152-5001

Christine In
Space and Naval Warfare Systems Center Pacific
53560 Hull Street, San Diego CA, 92152-5001

Stacey Curtis
Space and Naval Warfare Systems Center Pacific
53560 Hull Street, San Diego CA, 92152-5001

Corresponding Author:
Adriane Wolfe
619-553-5753
adriane.wotawa-berge@navy.mil

ABSTRACT

Energy is a critical resource for data centers that can impact system performance and sustainability. Data center systems can be dynamically managed and workloads can be scheduled to meet various operational energy goals. This paper provides a review of relevant research conducted in the area of data center energy management for managing from four perspectives: computing efficiency, cooling efficiency, renewable energy, and cyber–physical security. While energy management is considered from these perspectives, a theme is observed that energy is fundamental, and thus the perspectives collapse into one holistic systems perspective.

KEYWORDS

Data center, energy management, scheduling, efficiency, security

1. INTRODUCTION

The energy consumed by data centers has risen rapidly: from 2005 to 2010, energy consumption by data centers increased by 36%, and as of 2011 data centers accounted for nearly 2% of total electricity use in the U.S. [1]. Furthermore, a 2011 Gartner study listed the top six data center challenges reported by industry professionals; four of them were directly related to energy including infrastructure monitoring, energy efficiency, power density, and heat density [2].

Unsurprisingly, there is an active field of research investigating energy optimized data center designs and management methods [3]. While energy savings can be found via physical configuration, infrastructure architecture, and equipment selection these design choices are not always available and are generally planned to include extra capacity for growth and peak requirements [4]. Systems that are not designed to match their load operate inefficiently. To address this, dynamic energy management techniques are proposed.

This paper focuses on energy management methods, which are being developed to allow for dynamic scheduling. We describe energy management as the dynamic management of data center equipment to balance energy sources and loads for optimal performance. This paper begins with a brief background, followed by a review of dynamic energy management from four perspectives: computing efficiency, cooling efficiency, insertion of renewable energy, and cyber–physical security. Finally, the driving concepts of these sections are combined for envisioning a holistic energy management perspective.

2. BACKGROUND

This section provides background including a description on data center computing, energy, efficiency, and system characterization.

2.1. Data Center Computing

The primary operation of a data center is to provide computing capabilities. The computing equipment consists of servers, storage area networks (SAN), and network switches. The servers have traditionally hosted one operating system on which applications run. In recent years, the IT industry has trended toward virtualization where multiple operating systems can be hosted on one server and are managed by a middleware system called a hypervisor [5]. The hypervisor provides a simulated environment on a host server with which it manages the allocation of the server's CPUs and RAM for each virtual machine (VM) [6]. The allocation of resources can either be user defined or dynamically allocated.

Virtualization technology has led to the emergence of cloud computing, as well as slowing the growth of data center energy use [1]. According to the National Institute of Standards and Technology (NIST), cloud computing is a "model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [7]. Cloud computing is inherently an IT load management architecture that elevates virtualization from server scale resource optimization to shifting and scheduling of VMs across multiple servers within a network.

Virtualization and cloud computing technology enables data center service models. These service models are generally contract deliverables and measured based on Service Level Agreements (SLAs) and Quality of Service (QoS) [8]. In these service agreements, specific computing performance and total data center reliability requirements are defined by service providers and the application customers.

2.2. Data Center Energy

Data centers require power not only to support the computing equipment they house, but also to support the infrastructure that supports that equipment. The total power used by a data center, P_{DC}, can be broken down into computing (IT) power, cooling power, distribution power, and lighting power [9,10]. Researchers generally agree that IT and cooling systems present the most opportunities for improving total system energy efficiency [11]. Although there is little research on improving the efficiency of power distribution systems directly, they are an essential piece of the power management process [12].

The infrastructure for power distribution in a data center includes distribution components such as buses, transformers, and power distribution units (PDUs) [13]. Backup power is traditionally provided via uninterruptable power supplies (UPS) which provide short term battery storage, and diesel generators which provide longer term, on-site generation during grid outages [14].

Data center reliability is often referenced in terms of tiers. The Telecommunications Industry Association has developed requirements for four reliability tiers as seen in Table 1 [15]. These tier levels dictate both the redundancy type and the availability expectations specified in the service level agreements (SLAs). From the perspective of power the SLA availability requires the power to be available to provide service.

Table 1: Data Center Reliability Tiers [15]

Tier	Descriptor	SLA Availability
Tier 1	Basic	99.671%
Tier 2	Redundant Components	99.741%
Tier 3	Concurrently Maintainable	99.982%
Tier 4	Fault Tolerant	99.995%

2.3. Data Center Efficiency

Concern for energy efficiency motivates the investigation of relevant data center metrics. The two most widely used efficiency metrics measure the fraction of input electricity used for running servers, as compared to the total data center power consumption. The Data Center Infrastructure Efficiency (DCiE) is defined as:

$$DCiE = P_{IT} / P_{DC} \quad \text{(Eq. 1)}$$

where DCiE is the data center energy efficiency, P_{IT} is the power consumed by the IT equipment, and P_{DC} is the total power consumed by the data center [16,17]. A more commonly used metric is power utilization effectiveness (PUE) which is simply the reciprocal of DCiE [17]:

$$PUE = P_{DC} / P_{IT} = 1 / DCiE \quad \text{(Eq. 2)}$$

In a 2012 survey of 1100 data centers worldwide, the Uptime Institute found that data centers average a PUE of 1.8-1.89 [18]. The problem with this commonly used metric is that it assumes that the IT power is 100% effective in its use towards computing. However, IT systems can be managed to perform their functions inefficiently and some of their functions, such as fans, are not directly used to support computing [19].

These challenges highlight some imperfections in total data center system metrics. To the authors' knowledge there is no broadly accepted metric for measuring the computational efficiency of a total data center. In order to evaluate the true energy efficiency of data centers granular performance data and a variety of multi-dimensional metrics are required.

2.4. System Characterization

Data center energy can be characterized via granular system characterization. The power consumption is generally described from one or both of these activities: monitoring and modeling. As with other fields, the modeled and monitored system performances can be compared and used for efficiency refinement.

2.4.1. Data Center Monitoring

Energy monitoring (i.e., metering) of data centers can be used to characterize and optimize real-time physical system performance [20,21,22,23]. Servers, power distribution, and cooling systems can be monitored at a component level. Additionally, software can be used to aggregate data and provide total systems monitoring.

Energy related data for IT equipment can be collected through middleware that is offered by commercial vendors supporting specific devices. Virtualization hypervisors such as VMware and Xen, have product lines for monitoring virtual computing, networking and storage resources [5,24]. Through computing utilization and server specifications, estimations can be made for VM level power consumption. For the physical host machines, commercial vendors have developed solutions to communicate with the server's energy control system using energy management tools (e.g., Intel Data Center Manager, and iDRAC) [25,26, 27]. Advanced Configuration and Power Interface (ACPI) is used as an open industrial specification for smart metering and power management. It was designed for enabling operating systems' directed configuration, power, and thermal management [28]. These tools allow IT administrators the ability to remotely deploy, update, monitor, and maintain servers.

Monitoring can also be conducted for supporting infrastructure such as for distribution and cooling components. Components can be metered for power usage using building level smart meters or lab bench power analyzers [29,30]. In addition, some components such as uninterruptible power supply (UPS) and rack power distribution units (PDUs) can have built in power consumption metering capabilities [31,32]. Individual cooling components such as CRACs and chillers have built in control consoles that allow for the adjustment of temperature set-points [33].

In smart metering, the meters and devices are networked and the data is collected through communication protocols such as Simple Network Management Protocol (SNMP) or Windows Management Instrumentation (WMI). The network feeds the data to Data Center Infrastructure Management (DCIM) software tools. These DCIM tools are specifically designed for infrastructure and energy management. They have dashboards that can be used by the data center manager to monitor and implement policies to optimize the utilization of resources [34].

Energy monitoring techniques can provide energy management support for efficiency while ensuring availability and reliability of power and energy systems. Alternative to monitored characterization, simulation and modeling approaches can also be used to characterize system architectures and workloads [22,35].

2.4.2. Data Center Simulation

Datacenter modeling and simulation is an important approach to rapid system characterization and refinement. It offers several advantages over monitoring physical setups including repeatability, scalability, and reduced level of effort. However, they are based on models of the system and therefore may not accurately represent actual performance.

Computing energy simulation tools have been developed targeting different uses, with different assumptions, and to varying levels of maturity [36]. Two of these tools were made open source: CloudSim and GreenCloud [35,36]. CloudSim was developed by the University of Melbourne. It is a Java based API that simulates data centers to the level of virtual machines. With synthetic workloads, it implements user-defined strategies to optimize power consumption, while accounting for computational service-level agreement violations [37]. GreenCloud, developed by the University of Luxembourg, is a packet-level simulator that captures energy consumption at the level of server, switches, and link components. [38]. These and other simulation platforms serve as test beds for a variety of computer power characterization and optimization strategies [39].

While many researchers model cooling policies using Matlab [40,41], industry has developed specialized software to model and optimize data center cooling, using computational fluid dynamics modeling [42,43]. These models are described by the physical layout of the data center with air conditioning settings and physical heat load allocation. They compute the fluid dynamics and graphically represent the air flow and temperature in the physical space.

Both simulation and monitoring techniques can be used as tools to develop and refine data center energy management techniques for each of the four energy management perspectives, as investigated in this paper.

3. EFFICIENT COMPUTING

This section presents efficient computing including the systems background, power reduction theory, management techniques, applicable metrics, and a review of relevant research.

3.1. System Background

To achieve energy efficient IT systems, the general approach is to reduce the power consumption of the computing systems while maintaining service agreement performance. This can be conceptualized from a systems perspective where the system is the computing equipment, the input is power, and the output is service level agreement (Fig. 1). The goal of efficiency for any system is to maximize the ratio between the output and the input, in this case service performance versus input power.

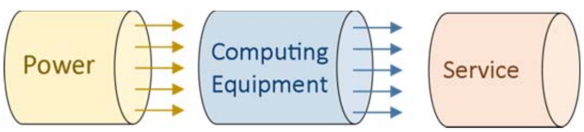


Figure 1: Systems perspective of computing efficiency.

3.2. Computing Power

The IT Power, P_{IT}, consists of the sum of the power consumed for computational loads, P_{Server}, the power for the data storage, P_{SAN}, and the power for the network equipment for data communication, $P_{Network}$ (Eq. 3) [44].

$$P_{IT} = P_{Server} + P_{SAN} + P_{Network} \quad \text{(Eq. 3)}$$

The servers typically consume the highest of the IT components, on the order of 70% of the IT power and therefore much research has focused on optimizing the server efficiency [17]. The server power consists of the sum of the internal components; CPU, memory, disks, peripherals, and the motherboard [17].

The CPU power accounts for at least 20% of the server power [45]. Authors have presented equations describing the CPU utilization as proportional to the overall system load:

$$P_{CPU}(u) = k * P_{max} + (1- k) * P_{max} * u \quad \text{(Eq.4)}$$

where P_{max} is the maximum power consumed when the server is fully utilized, k is the fraction of power consumed by the idle server, and u is the CPU utilization [35].

The power consumption in a CPU can also be correlated to the voltage and frequency at which it operates at and can be described by:

$$P_{CPU}(f) = ½ CV_f^{\,2}Af \quad \text{(Eq. 5)}$$

where C is the load capacitance, A is the activity factor, V_f is the CPU voltage, and f is the CPU frequency [45,47]. The CPU Voltage is driven by the CPU frequency f:

$$V_f = V_{max} * f / f_{max} \quad \text{(Eq. 6)}$$

where f_{max} is the maximum allowed CPU frequency and V_{max} is the corresponding voltage at that allowed CPU frequency [45].

It should be noted that P_{Idle} can be on the order of 60% of P_{Max} for a server [46]. The utilization consumption can be represented at the server level as:

$$P_{Server}(u) = P_{idle} + (P_{peak} – P_{idle})\, u \quad \text{(Eq. 7)}$$

where P_{server} is the power as a function of server utilization, u, which is multipled with the difference between the peak power, P_{Peak}, and idle power, P_{Idle} [10]. While the model in Eq. 7 is commonly used and has clear correlation with subsequent server power models, it has been found that the relationship can be refined as a nonlinear model where u^r is experimentally obtained to minimize the square error (Eq. 8) [48].

$$P_{Server}(u) = P_{idle} + (P_{peak} – P_{idle})\,(2u – u^r) \quad \text{(Eq. 8)}$$

In situations of virtualization the server power can be described by:

$$P_{Server} = P_{Idle,\, default} + N_{VM} * b \quad \text{(Eq. 9)}$$

where the server power is a function of the number of VMs on the server, N_{VM}, multiplied by the utilization of each VM, b, summed with the amount of power needed when no VM is running, $P_{idle,default}$ [10,40].

3.3. Management Techniques

There are four primary techniques that have been studied to achieve dynamic IT efficiency: server mode scheduling, VM dynamic scheduling, dynamic voltage and frequency scaling, and power capping (Table 2).

Table 2: Compute Efficiency Management Techniques

Technique	Variable	Sources
Server Mode Scheduling	On/off/sleep mode	Liu, 2009 [21] Goiri, 2010 [22] Berral, 2011 [23] Banerjee, 2010 [49]
VM dynamic scheduling	VM allocation	Tang, 2007 [20] Harizopoulos, 2009 [50] Liu, 2009 [21] Goiri, 2010 [22] Berral 2011 [23] Pakbaznia, 2009 [40] Banerjee, 2010 [49] Li, 2011- iSwitch [51]
Dynamic voltage and frequency scaling	Frequency	Beloglazov 2012 [35] Pakbaznia, 2009 [40] Li, 2011- SolarCore [52]
Power Capping (Single Threshold)	Device power	Liu, 2009 [21] Beloglazov, 2012 [35]

Server Mode Scheduling: Intelligent mode scheduling refers to turning server modes to on, off, or sleep modes. When a server is not in use it can be turned off to avoid P_{Idle} consumption (Eq. 9). If there is anticipation that the resource will be needed in short order, it can be put in sleep mode (as opposed to off mode) to minimize start-up time delay and energy losses.

VM Dynamic Scheduling: To consolidate underutilized computing resources virtual machines can be shifted from one server to another thereby achieving higher efficiency on the host server, and achieving zero utilization on the consolidated server (Eq. 10).

Dynamic Voltage and Frequency Scaling (DVFS): DVFS involves the automatic reduction of the frequency driving the CPU when it is under-used to reduce its power consumption and associated heating (Eq. 5,6) [10]. The actual frequency is set by a sensing circuit called a governor, which employs four user-settable options: power-save (frequency runs low), conservative (frequency changes in steps), on-demand (frequency toggles between low and high), and performance (frequency runs high). Since the cpu voltage is linked to the frequency, it is computed and set by the governor.

Power Capping: Power capping involves limiting the server's power consumption by setting a maximum power threshold.

Policies generally use combinations of the available techniques. Here typical example relationships between the techniques are described. When a server is required for computation it can be turned on, when it is not required it can be turned off. However, servers are often not completely idle; they are more often under-utilized. Studies have shown that typical data centers have an average utilization of 30% per server [53]. In the case of under-utilization two techniques are available: VM dynamic scheduling, and dynamic voltage and frequency scaling (DVFS). VM scheduling can be used to consolidate VMs so that underutilized servers become empty of VMs. The server mode scheduling technique can then be used to turn off or set in sleep mode the server that is no longer being utilized to reduce P_{Idle} consumption. If these workloads are under utilizing the server, then dynamic voltage and frequency scaling can be used. Power consumption can also be limited via power capping.

3.3. Evaluation Metrics

When management policies are enacted the primary types of metrics used are scale, computing performance, power reduction, and SLA agreement.

Scale: The system in question is generally described in terms of applications, components, or VMs [35, 54].

Computing Performance: The level of performance generally refers to the number of VMs or level of utilization. These variables drive the component level power consumption (Eq. 4, 7, 8, 9) [35, 54].

Power Reduction: Power is generally measured for a baseline of the given system and then improved via a management technique to achieve some level of reduction in power [11, 21, 35, 54].

Service Level Contract Agreement (SLA): Computing reliability is measured by considering contract performance [35,54]. Computational SLAs are the primary type of service level agreement (SLA) used in this section. This can be defined as a percentage of the SLA violation events normalized by the total processed time frames. An SLA violation occurs when a VM does not receive the number of MIPS (million instructions per second) that are requested.

Unsurprisingly, the metrics used to consider computing efficiency can be organized under the systems perspective. Typically the first metrics presented are those that describe the system generally in terms of the scale and computing performance. Then some policy is enacted using the management techniques and a description is made regarding the required power (input) compared to the impact to the service level agreement (output).

3.4. Relevant Research

This section reviews literature that investigates dynamic management approaches for computing efficiency, (i.e. the reduction of IT power (P_{IT})). This literature shows the evolution of how the primary computing management techniques are combined for dynamic scheduling.

Harizopoulos et al., 2009, argued and presented motivation for energy efficient computing research as compared to performance-oriented research [50]. Authors argued that cost-benefit trade-offs must be made between incremental enhancements in performance and total power costs. Three energy efficiency methods are presented: energy-aware optimization, resource use consolidation, and redesign for maximum energy efficiency. The paper does not demonstrate the proposed methods [50].

Tang et al., 2007, developed a computing application placement optimization product [20]. The application placement problem defined by the authors was: given a set of machines with constrained resources and a set of virtualized Web applications with dynamically changing demands, where to place each application and how many instances of that application to run. The placement product could place thousands of applications compared to traditional approaches that placed hundreds of applications managed by a commercially available hypervisor. The new algorithm outperformed the old algorithm by 25% demand satisfaction and it reduced placement changes by up to 90%. The paper did not study the energy efficiency impact of the proposed controller [20].

Liu et al., 2009, developed an architecture titled GreenCloud to address energy efficiency and sustainability concerns in data center management. The goal was to consolidate virtualized workloads and reduce data center power consumption for cloud computing environments while guaranteeing the performance for many performance-sensitive applications. GreenCloud is an internet data center (IDC) architecture that automatically makes the scheduling decision on dynamically migrating and consolidating VMs among physical servers to meet the workload requirements [21]. Unlike Tang et al (2007), this paper was able to study the impact of its policy on power consumption. The energy reduction over a 12 hour period of time ranged from near 0% to 57% reduction in energy [21].

In the paper authored by Goiri et al., 2010, a similar focus is given to scheduling of migration of virtual machines to achieve energy efficiency for high performance computing (HPC) data centers [22]. The article presents a score-based scheduling approach that weights criteria to decide whether to migrate a virtual machine and to optimize where it is migrated to. It investigates and optimizes algorithms to solve a criteria-weighted matrix to find optimal solutions. It included the metric that service level agreements (SLAs) be maintained. The approach was more complex than that presented by Liu et al (2009). The general approach was to measure power consumption behavior of real computers managed by Xen, simulate them using OMNet++, then scale them to simulate large scale data centers [55]. The model predicted a total energy estimation 2.4% below the experimental consumption [22]. This suggests that workloads can be reasonably modeled and scaled for rapid development. The impact of migrating workloads and changing server modes resulted in a system with a maximum power consumption of 3000 kW down to a system consuming 500 kW. However, client satisfaction during the lowest power consumption modes was between 86-90% in comparison to 100% during high power consumption modes. While the highest level of energy savings was higher than the results found in Liu et al (2009) they were shown to be low in client satisfaction. These results suggest that a cost-benefit tradeoff between energy efficiency policies and service level agreement is required [22].

Berral at al., 2010, proposed a framework that provides an intelligent consolidation methodology in order to maximize performance and minimize power consumption [56]. The framework covers the control cycle: from the acquisition of real power measures to the scheduling of the resources. They used different techniques seen in previously mentioned literature such as server mode scheduling, and power-aware VM consolidation algorithms. They added and compared machine learning techniques in order to predict power consumption levels, CPU loads, and SLA timings, and improve scheduling decisions [56]. They used machine learning techniques in order to predict, from a given set of machines and set of jobs, the resulting client satisfaction level of each job and power consumption before placing tasks in servers or migrating tasks. Three types of workloads were run on five policies and their relative performance was compared. For the Heterogeneous and Service workloads the machine learning approach had the lowest power with a 23% and 30% energy reduction respectively. While this is a lower power savings than represented in previous papers, the fact that machine learning can assist in achieving power was demonstrated [56].

Beloglazov et al., 2012, also proposed and evaluated techniques for dynamic reallocation of VMs to minimize energy consumption while ensuring reliable Quality of Service (QoS). The QoS is evaluated using the % of Service Level Agreements violations [35]. CloudSim was selected as a simulation toolkit. A benchmark approach of a Non Power Aware (NPA) policy was compared to three energy efficiency methods: dynamic voltage and frequency scaling (DVFS), single threshold [or power capping] (ST), and minimization of migrations (MM). The “MM 30-70%” policy was successful in reducing energy consumption by 85% with 1.1% SLA violations which was a significant improvement and the highest consumption reduction with low SLA violations achieved as compared to the previously mentioned papers [35].

In summary, there are great opportunities for energy efficiency in computing, with research reaching 85% power reduction. There is a clear trade-off between power reduction and SLA agreement. However, it seems the management policy selected drives the balance of that relationship and the policy optimization is workload type dependent. The following section transitions to cooling efficiency.

4. EFFICIENT COOLING

The computing workloads described in the previous section require cooling to maintain their operations. This section presents efficient cooling in an organizational structure similar to the efficient computing section.

4.1. System Background

A critical enabler of data center computing is the cooling systems which remove waste heat from IT equipment [57]. The cooling system can be conceptualized from a systems perspective similar to the one used for computing equipment. The input is power. The system is the cooling equipment, which has the output of a service to the computing equipment, to maintain the required operating temperature (Fig. 2). The goal is to maximize the ratio between the output and the input, the maintenance of required temperature versus the power consumed by the cooling system.

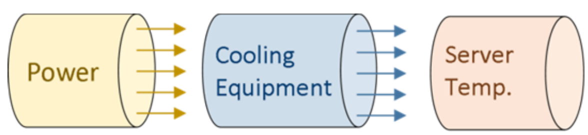


Figure 2: Systems perspective of cooling efficiency.

The heating, ventilation, air conditioning (HVAC) industry provides numerous types of devices and cooling technologies that are used to cool data center computing equipment. The majority of these systems consist of one or more of the following basic elements: heat transport (e.g. fans, pumps), heat exchange (e.g. coils or vents), compressor and evaporative assist devices (e.g. cooling towers, wetted filters). During the cooling process, heat transport and compressors consume electrical energy and the evaporative assist consumes water.

Computer room air conditioner (CRAC) units are commonly used to provide air- based cooling in data centers. A CRAC unit monitors and maintains the temperature, air distribution, and humidity in a data center by supplying chilled air to achieve a specific temperature range required for optimal server operation [58]. Cool air must be circulated to absorb the heat that is generated by the IT equipment in the room [59]. The CRAC unit completes this cooling cycle by first pumping the cool air through either an elevated floor or directly into the room. The chilled air dispenses and rises through perforated floor tiles or ducting in designated sections of data centers that form the cold aisles. The cold air flows through the racks, and as the servers and racks intake the cool air, they exhaust hot air into the opposing section, which is hence called the hot aisle. The CRAC unit then pulls in the hot air from the hot aisle and the heat is transferred to a cooling fluid and exhausted out of the room [60]. Chiller units provide a cooling liquid to chill the hot air drawn by CRACs into cool air exhausted out of the CRACs.

There are also emergent cooling architectures that use "free cooling" from cold outside air or liquid cooling for high thermal transport [61]. While these will have respective technology specific management requirements, the core fundamental concepts in this section remain relevant.

4.2. Cooling Power

To achieve energy efficient cooling systems, the general approach is to reduce the power consumption of the cooling while maintaining temperature requirements of the servers.

The cooling power, $P_{Cooling}$, consists of the sum of the power consumed by the chiller, $P_{Chiller}$, the tower, P_{Tower}, the pump, P_{Pump} , and the blower, P_{blower} (Eq. 10) [9]. Generally the two largest consumers of energy consumption are the chiller and blower systems (i.e. the CRACs).

$$P_{cooling} = P_{chiller} + P_{tower} + P_{pump} + P_{blower} \quad \text{(Eq. 10)}$$

In recent attempts to reduce cooling requirements of chillers, data center operators are working to run data centers at higher temperature set-points [9]. However, equipment with high power density or low air flow delivered could result in overheating and computing system failure. To avoid this, management of the temperature at the servers is required to reduce the risk [57]. This is a primary consideration for the development of dynamic management techniques.

4.3. Management Techniques

There are four primary techniques that have been studied: temperature set-point, CRAC variable frequency management, directed cooling to physical location, and computational tasking based on physical location (Table 3).

Table 3: Cooling Efficiency Management Techniques

Technique	Variable	Sources
Supplied air temperature (SAT)	Temp Setting	Zhou, 2011[41] Pakbaznia, 2009 [40] Banerjee, 2010, [49] Huang, 2011 [9]
Variable Frequency Drive (VFC) CRAC	Frequency	Zhou, 2011 [41]
Directed cool air delivery	Physical location	Zhou, 2011 [41] Pakbaznia, 2009 [40]
Computational tasks based on physical location	VM location, server mode, server frequency	Pakbaznia, 2009 [40] Banerjee, 2010 [49] Said, 2011 [10]

Supplied Air Temperature (SAT): The temperature settings for the chiller and CRACs can be set to control the temperature that the fluid is cooled to. Typically the chiller set-point is held constant and the CRAC set-point is managed to control the temperature of the cold air supplied to the servers.

Variable Frequency Drive (VFD) CRAC: Some CRACs have variable frequency drives which allow the blower to be tuned based on the required tonnage for cold air. However, many CRACs do not have this feature [2].

Directed cool air delivery: The air can be directed based on the physical location of heat loads. This can allow for more cold air to be delivered to servers that are running hot, and less cold air can be delivered to areas that do not require additional cooling.

Computational tasks based on physical location: The computational tasks can be assigned to locations that are physically easier to cool. This relies on the energy management techniques developed for computing efficiency, especially VM allocation.

There are two general ways to schedule for cooling efficiency, one is to manage the cooling supply and the other is to manage the cooling demand. In supply management, the temperature of the air, the amount of air flow, and location of air delivery are controlled to reduce waste. On the demand side, the computing loads, which correlate to heat loads, can be managed to consolidate or redistribute cooling requirements based on their physical locations.

4.4. Applicable Metrics

When management policies are enacted the primary types of metrics used are description of cooling, IT environmental, and power reduction.

Cooling: The system in question is generally described based on the characterization of the cooling mechanisms, tonnage or cubic feet per minute (CFM) of cooling, kW of heat load, and heat recirculation [40, 41, 49, 9, 40, 41, 49].

IT Environmental: This relates to the conditions of the IT equipment as compared to their required temperature and humidity conditions [9, 40, 41, 49].

Power Reduction: Power is generally measured for a baseline of the given system and then improved via a management technique to achieve some level of reduction in power [9, 10, 40, 41, 49].

The metrics used can be organized under the systems perspective. Typically the first metrics presented are those that describe the relationship between the cooling system. Then some policy is enacted using management techniques and a description is made regarding the input, the required power compared to the output, the IT environmental status.

4.5. Relevant Research

This section reviews relevant research that investigates management policies for cooling efficiency.

Zhou et al., 2011, proposed a holistic data center cooling scheme in order to optimize the provisioning, transport, and distribution of cooling resources [41]. While the temperature of the recirculated hot air is beyond direct control, the cool air delivered to the rack inlets were adjusted through tuning of SAT and VFD of the CRAC units, and vent tile openings to regulate the delivery to the rack. In order to minimize the total flow and thermodynamic work done by the cooling system, zonal and local cooling actuation was coordinated in a unified framework with a Model Predictive Controller (MPC). The Model Predictive Controller (MPC) optimization problem was formulated and implemented in Matlab. The Matlab optimization toolbox function "fmincon" was used to solve the constrained optimization problem. Based on experimental results, the authors claim that using their model, up to 36% of CRAC unit blower power can be saved [41].

Unlike, the MPC approach which focuses on the supply side of cooling, Pakbaznia et al., 2009, focuses on the demand side of cooling for power minimization [40]. The goal was to minimize the total data center power consumption by: 1) determining the optimum value of supplied cold air temperature, 2) turning various servers and chassis ON/OFF, and 3) for ON chassis determining the number of the ON servers and their corresponding cores' voltage-frequency (v-f) levels. The cooling power was reduced by choosing an optimum supplied cold air temperature value. Server power was reduced by appropriately assigning incoming tasks to different servers and setting the proper (v-f) level for each server based on assigned tasking. Using Matlab modeling results the authors claim an average of 13% power saving for different data center utilization values compared to a baseline task assignment technique that does not perform consolidation [40].

Banerjee et al., 2010, similarly focused on a cyber-physical oriented coordinated job and cooling management. The authors developed an approach for high performance computing (HPC) data centers to reduce the total computing and cooling energy consumption [49]. Spatial job scheduling algorithms, which decide on which servers the jobs are executed on are evaluated in terms of their energy inefficiency [49]. The energy inefficiency is measured by the SP-EIR metric, which depends on the heat recirculation in the data center and the thermostat set temperatures of the CRAC unit. The Highest Thermostat Setting (HTS) algorithm was proposed, which places jobs based on the heat recirculation and the servers' requirements on the CRAC thermostat settings to meet their respective redline temperatures. The results show that the energy savings achieved by the HTS algorithm are greater than the savings previously predicted using other algorithms. The energy savings ranged from 0.7% to 12.4% when idle chassis were left on and 0.4% to 23.7% when they were turned off [49]

Huang et al, 2011, proposed two thermally-aware power optimization techniques that target server fan power, one for servers and one for data centers [9]. Server-level fan energy is frequently the second highest source of energy consumption in server and storage systems and can consume 23% of total server power. Thermal-Aware Power Optimization management policy was developed for data centers (TAPO-dc) scale systems. TAPO-dc switches between two distinct HVAC chiller setpoints (high and low) for a cooling zone. It optimizes aggregated HVAC and server fan power, and can achieve a 12.4%-17% reduction in total data center power with no performance penalty. Thermal-Aware Power Optimization is a management policy for servers (TAPO-server) that uses runtime measured power to adjust server thermal set-points and optimize aggregated server fan and leakage power. The TAPO-server reduced server power by 5.4% when the server processor was heavily exercised with no performance penalty [9].

Said et al., 2011, authors provided a data center energy model along with an optimization energy management algorithm designed to increase the overall data center energy efficiency [10]. Unlike, previous papers this approach considers power distribution, cooling, and the computing systems holistically. The objective of the proposed optimization algorithm is to minimize energy consumption while maintaining SLAs. Authors claim that their proposed algorithm has yielded a 20% energy gain over a previous algorithm [10].

Cooling efficiency methods have demonstrated the potential for significant energy consumption reduction up to 32%. Researchers are less oriented toward service level agreement than for computing efficiency, which may be due to the opportunity to achieve efficiencies without threatening the computing output. Significant research emphasis is being focused on the close relationship between the cooling and computing heat load management. As cooling management is being conducted in concert with computing management, it is seen that SLA becomes a more critical metric.

Data centers are interested in both reducing their energy demand through improved efficiencies as well as developing renewable energy supplies. The next step that researchers are taking is looking at how renewable energy can reduce the cost of energy and associated carbon emissions.

5. RENEWABLE ENERGY

The insertion of renewable energy as a power source for data centers is a growing industry and research trend. Several prominent IT companies have begun transitioning to the use of renewables [62,63]. Over 30% of Google operations are powered by renewable energies like solar and wind and nearly 100% of Apple facilities are powered by some form of renewable energy [62,63]. Renewable energy sources, however, pose unique challenges to energy management. This section provides review of scheduling for renewable energy utilization at data centers.

5.1. Systems Background

Traditional data centers have been powered by the power grid and backed up using diesel generators. Emergent trends have emphasized the insertion of renewable energy for environmental and economic benefits. The use of renewable energy can be conceptualized, similarly to previous sections, from a systems perspective where the input is green (renewable) and brown (grid) power, the system is the data center, and the output is service (Fig. 3). For this system, the goal is to have high renewable energy usage and for the load to match the source so that the power system remains stable and efficient. If the system becomes unstable, then the service is at risk of not being fulfilled.

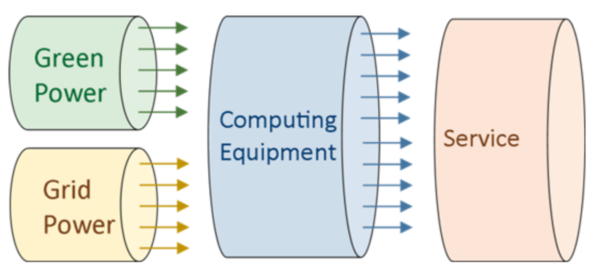


Figure 3: Systems perspective of renewable energy insertion for computing systems.

Renewable energy sources can range from relatively constant sources such as geothermal, thermal reuse, hydroelectric, and variable sources such as wind, and solar [64,65]. Wind and solar are variable due to variations in weather systems, the time of day, and their geographic location [65]. This variability stands in contrast to the relatively reliable and constant source of power provided the electric grid. This variability drives the need for energy management techniques specific to renewable sources. The challenges of integrating green energy is matching the load to the variable renewable energy availability. Too little green energy would lead to a mismatch of computing and desired output and lead to SLA violations. In contrast improperly utilized, green energy may be lost and inefficiently used.

5.2. Relevant Techniques

The scheduling techniques for variable power production are tailored to match the data center workload to changes in the power availability and to make the best use of the power when it is available. There are two primary techniques that are key to optimizing the utilization of these renewable energy sources: geographical scheduling, and local load balancing (Table 4).

Geographical Scheduling: Geographical scheduling is the practice of shifting workloads between geographically separate data centers based on the renewable energy availability.

Local Load Balancing: Local load balancing involves implementing techniques designed to facilitate the integration of renewable energy into powering a datacenter located at a single location. The two types of local load balancing found in literature are power distribution and batch scheduling.

Table 4: Renewable Energy Management Techniques

Technique	Sources
Local Load Balancing	Stewart, 2009 [66] Deng, 2011 [67] Li, 2011- iSwitch [51] Li, 2011- SolarCore [52] Goiri, 2011 [68]
Geographical Load Balancing	Le, 2009 [69] Gao, 2013 [70] Liu, 2011 [71] Zhang, 2011 [72] Lin, 2012 [73] Chen, 2012 [74] Ghamkhari, 2012 [75] Akoush, 2011 [76]

Generally speaking research regarding local data center optimization tends to consider a combination of job scheduling and power distribution techniques while, in contrast, multi- data center optimization tends to focus on geographical load balancing techniques.

5.3. Applicable Metrics

When the previously mentioned management policies are enacted the primary types of metrics used are system characterization, renewable energy usage, cost reduction, and SLA agreement.

System characterization: The system in question is generally described based on the scale and set-points of the renewable energy source and the computing load [51, 52, 66, 68, 76].

Renewable energy usage: The amount or percentage of renewable energy used. This can also be correlated to carbon footprint which has sustainability implications [51, 52, 66, 67, 68, 69, 74, 75, 77].

Cost reduction: The amount of renewable energy used can be used to calculate the cost of total operational energy or the cost per unit of energy [66, 71, 69, 72, 75, 77].

Typically the first metrics presented are those that describe the system generally in terms of the data center scale and settings. Then some policy is enacted using the management techniques and a description is made regarding the required level of renewable power input compared to the impact to the service level agreement. The levels of renewable power and grid power input can also be evaluated from a cost perspective.

5.4. Relevant Research

This section reviews relevant research that investigates the energy management of variable renewable energy sources from the perspective of local load balancing and geographic load balancing.

Local Load Balancing

Stewart and Shen, 2009, studied the potential for improved use of renewable energy resources by multiple datacenters processing web server requests [66]. They collect data from two wind farms over a period of several years and observe that increasing the power threshold that regulates switching between renewable and grid power sources allows greater consumption of renewables at lower cost than grid energy. The authors conduct computer simulations which show that the power draw of several server requests is predictable. The results obtained in this paper suggest that a cloud computing service could submit web server requests to several datacenters that maximally exploit renewable energy sources [66].

Li et. al., 2011, proposed the implementation of SolarCore, a multi-core power management architecture designed to maximize solar power generation via load matching [52]. Due to large energy loss and negative environmental impact involved in storing solar power in batteries, it is desirable to utilize solar power immediately upon generation. As long as the power drawn from the solar panel is above the power transfer threshold, the solar panel applies a maximal power point (MPP) tracking algorithm to obtain maximum power from the solar panel. Dynamic Voltage-Frequency Scaling (DVFS) is employed to reduce CPU clock frequency on a per core basis so that the power required by the cores matches the power output of the solar panels. The authors can capture on average 82% of available solar energy with this method, and improve computational performance by 10.8% compared to a round-robin scheduling scheme [52].

Goiri et. al., 2011, emphasized job scheduling, authors propose GreenSlot, a parallel batch job scheduler that seeks to reduce grid energy consumption and cost for the datacenter [68]. GreenSlot is an alternative to the Linux SLURM job scheduler. In contrast to SLURM, GreenSlot seeks to maximize usage of solar energy, reduce the cost of brown energy, while meeting performance deadlines. Appropriate compensation is given to the job requester for unmet performance deadlines. In the authors' study, the implementation of GreenSlot in the solar-powered datacenter can led to a 117% increase in green energy consumption and up to 39% decrease in grid energy [68].

Geographic Load Balancing

The literature explores several different aspects of Geographic load balancing including whether green energy may be relied on exclusively, the cost consideration, and algorithms and strategies to run.

Geographic load balancing (GLB) was first introduced by Le et al., 2009. Le investigated the reduction of brown energy usage and cost using renewable energy consumption across multiple datacenters by load balancing according to time zone, electricity prices, and availability of renewable energy. The load distribution policies developed demonstrated a trade-off between energy cost savings and reduced energy consumption. Using the EPrice and GreenDC policies which assume access to a fixed amount of renewable energy, the hypothetical datacenters could achieve 25% cost savings or 38% reduction in brown energy consumption, respectively, while satisfying service level agreements (SLA) requirements [69].

Liu et al. developed two distributed algorithms for achieving optimal geographical load balancing [71]. They characterized cost as a linear combination of energy cost and lost revenue during the delay of requests. In trace-driven numeric simulation, they found that cost savings of over 40% could be achieved during light-traffic periods. They also showed that dynamic pricing, pricing of energy based on its percentage of brown energy, increased green energy [71].

Lin et al. developed an online algorithm for geographic load balancing [73]. They applied “receding horizon control” (RHC) a predictive algorithm to look ahead and predict the peaks and troughs of renewable energy availability. They showed that the algorithm performed well in homogeneous settings—in which all servers can serve all jobs equally well. However, realistically there are differences in propagation delays in servers and differences in electricity prices at different locations which can cause the algorithm to perform poorly. To address this, the authors developed a variety of the RHC algorithm that is mathematically optimized to perform well in conditions of heterogeneity [73].

Chen et al. also targeted the minimization of the use of brown energy, with the more particularized goal of managing batch job workloads [74]. Taking into account factors like environmental temperature and its impact on data center cooling, the quality of service of the batch jobs, and cost, the group developed an approach they called MinBrown. Simulations were performed on real workload and solar energy traces. They found that their approach yielded a 21% improvement compared to alternate green-aware approaches [74].

Akoush et. al., 2011, presented an architectural computing framework for utilizing otherwise unused renewable energy by multiple datacenters [76]. The motivation for their work is the observation that much renewable energy, e.g., solar and wind power, is un-utilized. The case study performed by the authors features two hypothetical datacenters with 2 MW of average combined annual renewable power capacity. The authors found that the number of events requiring a VM migration between renewable datacenters is about one a day. The authors found that the total amount of downtime per live VM migration was on a 10 Gbps wide-area network spanning 10,000 miles is 677.1 ms, which leads to a total yearly VM downtime of 416.4 seconds. This yearly per VM downtime represents 0.3% of the allowed downtime per VM for a datacenter with a 99.5% uptime per year service level agreement [76].

Ghamkhari et. al., 2012, developed a mathematical framework for modeling the optimal distribution of workload among several datacenters powered by renewable energy and brown (grid) energy [75]. The authors incorporate several real-time factors into their model, such as wind power generated, electricity price, incoming workload, and Quality of Service. Total cost was then optimized by distributing the workload to multiple servers based on the above factors. With the optimal workload distribution, a cost savings of 29% and a total savings of $87,970 are obtained over a ten-day time period. The mathematical model is shown to achieve significant energy cost savings while maintaining quality of service [75].

Liu et al., 2013, proposed the usage of demand response to curb energy consumption during coincident peak hours [77]. Liu et al. created two algorithms that minimize costs and maximize the usage of local generation, such as solar arrays to avoid the coincident peak of energy consumption, known as demand response. During coincident peak, power companies overcharge on electricity to discourage power consumption during hours of the day when the local energy consumption is high. With the proposed algorithms, data centers can workload shift and switch off to local PV power generation to avoid coincident peak, thereby reducing costs and making the most out of solar panel energy generation. As a result, a 35-40% reduction of costs and 10-15% reduction of emissions with the use of workload shifting and local power generation was achieved [77].

6. CYBER SECURITY

The scheduling of the management of data center systems for IT efficiency, cooling efficiency, and renewable energy insertion all rely on dynamic scheduling of physical systems using virtual (i.e. cyber) management techniques. This involves turning on and off systems, shifting workloads, and changing set-points, all of which are susceptible to cyber threats and attacks. This section investigates some of the most common cyber threats as well as investigates the relevant components of cyber-physical security from an energy management perspective.

6.1. Conceptual Description

This section focuses on energy security as it relates to cyber and cyber-physical threats for data centers from two perspectives:

Security of Primary Systems: this refers to the security of the primary data center systems, i.e., IT, cooling, and power distribution components.

Security of Energy Management: this refers to the security of the energy management system which manages the techniques in order to obtain reliability, efficiency, and insertion of renewable energy.

There are two main types of security threats to the energy and power systems: physical and cyber. Recently, an emerging research concept has been cyber-physical security, where the security of the physical system is compromised via a cyber-attack. An example of a cyber-physical attack is the Stuxnet worm. This worm was transferred to targeted systems virtually [78]. It then targeted Siemens industrial systems to monitor configurations and control resources. The Stuxnet worm is seen as the first publicly known example where the physical security of an industrial control system was compromised via a cyber-attack. This section takes particular interest in the cyber-physical relationship and proposes that the energy management techniques described in previous sections are both a potential threat to, as well as an enhancement for, this type of security as they have the potential to provide cyber-physical access into various systems.

6.2. Security Risks and Management

There are a number of types of security risks that have been identified in relation to cyber security relevant to data center energy management as seen in Table 5.

Table 5: Security Risks	
Risk	**Source**
Excessive Power Consumption	Palmieri, 2011 [47]
Denial of Service attack	Palmieri, 2011[47]
False Data Injection	Seo, 2011 [79]
Eavesdropping	Seo, 2011 [79]
Result Tampering	Seo, 2011 [79]

Risks to Security of Primary Systems: Denial of service attacks and excessive power consumption are risks to the primary data center systems. These represent cyber to physical threats and energy management techniques in the computing efficiency, cooling efficiency, and insertion of renewables sections of this paper can be applied to reduce the impact of these threats.

Risks to Security of Energy Management: Eavesdropping, result tampering, and false data injection are risks to the energy management system and can be used to compromise the integrity of the systems. They are based on traditional cyber security concepts, and as such the primary management techniques for their risks are Information Assurance (IA) techniques. Additionally, false data injection and result tampering of energy systems with control functions could result in cyber-physical threats if they cause the management system to make bad decisions when controlling the primary systems. This threat can limit the adoption of dynamic energy management solutions for security sensitive and critical systems.

From an energy and physical security perspective, the security evaluation and metrics can be considered from the systems perspective previously presented. The system is the data center equipment overall, the input is power, and the output is service. The new aspect to consider are risks that there are threats to the data center where it either suddenly demands more power than capacity available or the power supply is suddenly reduced so that the system is not balanced and therefore fails to meet service level agreements (Fig. 4a and Fig. 4b).

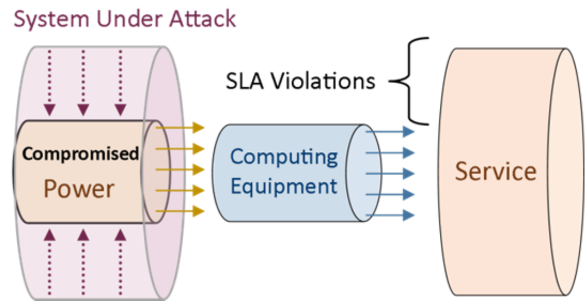


Figure 4a: Systems perspective of primary energy security risks highlighting the impact of an attack on the power grid.

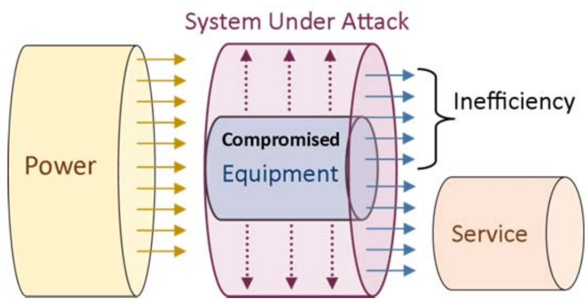


Figure 4b: Systems perspective of primary energy security risks highlighting the impact of the computing equipment requiring extra power.

6.3. Applicable Metrics

The metrics for measuring these cyber-physical risks are mainly dependent on the type of risk. For the primary systems the reliability of the physical system is the primary metric. For the security of the energy management the information assurance (IA) industry has developed metrics.

For primary systems the security of the system is measured from the perspective of reliability. As referenced in the background, data centers have specific contract agreements that specify the uptime requirements of their data center [15]. Typically, data center managers monitor their level of excess capacity as a measure of their ability to perform under increases in workload or decreases in the power supplied. This is one of the causes for the typical under-utilization of data center systems.

The IA industry is currently struggling with the challenge to define cyber security metrics. This is because a cyber-attack can be disguised as a regular transmission. In the past, system administrators would judge the security level of each system using their best judgment and intuition. Today security professionals monitor traffic on their network using metrics developed in-house. Vaughn et al created a standard of categories used when developing in-house metrics (Table 6) [80].

Table 6: IA Security Metric Categories	
Category	**Example**
Objective/Subjective	Annual down time/User training
Quantitative/Qualitative	Failed login attempts/Self-assessment levels
Static/Dynamic	Out of date/ Up to date
Absolute/Relative	Num. of engineers/Num. of vulnerabilities
Direct/Indirect	Num. of packets rejected/ISO Standard 15408

6.4. Relevant Research

This section reviews relevant research that investigates the cyber and cyber-physical relationships of energy management for data centers.

Tram et al., 2008, discussed integrating metering with response and tracking of security and power incidents [81]. Smart metering is used for outage management and has useful functions such as outage notifications (last gasp functionality), restoration notification, and on-demand read. This paper supports the concepts of the cyber-physical risks because the energy management system can be used to manage physical threats and also can have network risks that introduce risks to the management of those physical threats [81].

Kant et al., 2011, introduced the potential cyber risk of Configuration Management which is the combination of all the configuration information and the way it is stored and managed in the data center [82]. This information ranges from devices for computing, storage, and communication. Configuration management creates a potential opportunity for hackers to access the information and impact data center operations. This paper then discusses the security vulnerability of configuration management, and the effects of attacks on each component of a data center. A security mechanism called Sentry exploits the hierarchical setup and redundancy in data centers [82]. The threat models of Sentry make the assumption that the attackers can compromise the configuration data of the assets, but they cannot take full control of the assets [82].

Seo et al., 2011, considered both cyber and cyber-physical security and discusses an approach that enhances reliable power, ensures information security, communicates effectively, and detects malicious behaviors [79]. They highlight energy data as sensitive and energy reliability as a complex problem for utilities to manage. Four main types of attack models are studied: eavesdropping, result tampering, excessive power consumption, and false data injection [79].

Palmieri et al., 2011, focuses on the concept of excessive power consumption that Seo et al., 2011, introduced. The authors suggested that power consumption can be used to indicate a security breach [47]. A hacker compromises a CPU by increasing its work load to use up its resources [47]. This increase causes an increase in CPU power consumption. The gap between idle state and maximum power state can be up to 90%, during a CPU oriented denial of service attack. This could exhaust the power budget. The percentage of power utilized can be represented by Equation 13 where P represents power, f represents frequency and min and max subscripts represent the respective values at minimum and maximum loading [47].

$$\%P = P_{max} / P_{min} + 1 = f_{max}^{2} / f_{min}^{2} + 1 \text{ (Eq. 11)}$$

A hacker compromises a storage device by forcing it to accelerate and decelerate quickly using read/write operations. In the study, the power required for a read was 13.3 μW/Kbyte while the power required to write was 6.67 μW/Kbyte. The read occurred 4-5 times more often than the write and can be described in Equation 14 where it is comprised of two major components, the power for the motor plus the power for operation dependent electronics. The power required for a disk operation can be characterized as:

$$P_D = K^2 \omega^2 / R + D_r w_r P_{read} \text{ (Eq.12)}$$

where velocity P_D the power of the drive is dependent on the angular velocity ω, the motor voltage constant K, the motor resistance R, the power required for a read P_{read}, the read demand D_R, and the weight factor w_r [47]. This paper ties cyber security to energy security and ends with an advocacy for managing systems from an energy perspective as one method of mitigating denial of service (DoS) impact. Voltage frequency scaling, down-clocking devices and forcing sleep modes are highlighted as potential management solutions that require scheduling.

Hlavacs et al., 2011, focused on the concept of eavesdropping that Seo et al., 2011, introduced, considering how monitored server power could be used to identify and track movement of virtual machines during live migration for cloud computing [83]. This paper introduces the potential that an attacker could track down and target specific VM's that are being utilized based on power information. This same information could also be valuable to data centers in communicating energy costs to customers. Previous assumptions have been that VMs are more secure because they are decoupled from hardware; however, this paper highlights reliance on the security of the hypervisor. Each VM on a rack can have its own hypervisor so that hackers cannot easily attack the one that is being utilized. However, hackers can determine which is being used from the power data. This means they only need to attack a single hypervisor-which can be undetectable. In the experimental method it is suggested that power data at the metered rack PDU is most realistic and creates valuable data with 1 second time sampling [83].

In summary, the tie between energy management and the data center security is increasing in complexity and emergent threats are being identified. Energy management can be a potential solution for managing threats, but can also be leveraged to create new threats, which should be taken into account when designing a secure energy management system and management policies.

7. HOLISTIC SYSTEMS PERSPECTIVE

This paper takes a systems perspective of data center energy in which we consider a system to be comprised of data center equipment, the input to the system is power, and the output is computing service (Fig. 5). We have considered this perspective with respect to various research areas of data center energy management to display the universal applicability of this concept. While this is a basic concept, we propose that it can provide the framework to combine seemly disparate sub-topics into one holistic perspective.

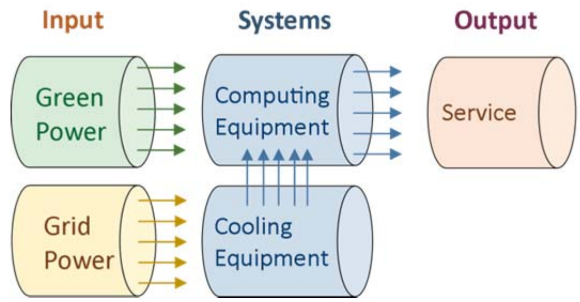


Figure 5: Holistic data center energy systems perspective.

Optimization and management of data center energy, whether for the stated goal of efficiency, renewable energy, or security, is conducted via this basic systems level concept. In the case of IT efficiency, the goal is to reduce the power while meeting computational service level agreements for given computing architectures. To achieve this power reduction, the computing architecture is effectively reduced to match the required computational load, thereby achieving efficiency. Similarly, in cooling efficiency, to achieve power reduction the cooling infrastructure is effectively reduced to match the heat load. The same techniques used to reduce the computational load can be used to reduce the heat load in a given physical location, allowing for a reduction in the cooling system, all the while delivering the cooling required for the heat load. In renewable energy utilization the power input is no longer considered as an unlimited resource and therefore when available, the load can be increased in terms of increasing computational through-put. For cyber-physical security the effective size of the workload or power supply can result in system in-stability.

Similarly, we propose that while researchers use a wide range of metrics they can be conceptualized and organized from this same systems level perspective as system characterization, power input, and service output. The effectiveness of a given policy or method is generally the ratio of input to output or more specifically it is the power consumed as compared to the ability to meet contract deliverables.

Interestingly the techniques proposed for managing cyber-physical security risks and renewable energy utilization often leverage the techniques developed for managing efficiencies. This suggests that the data center energy management techniques are independent of the particular goal of the management. The key to energy management considerations is matching load scheduling with source scheduling and not exceeding capacity limits. While this is a relatively simple concept, the optimization of this, using the primary techniques and metrics highlighted in this paper, to create a holistic energy management policy is a significant challenge. If achieved, it holds the potential to result in economic, environmental, and security impacts for the growing data center field.

8. CONCLUSIONS

Data center energy management policies have been developed by researchers with emphasis given to scheduling and dynamic load shifting. This research has been conducted in order to achieve computing efficiency, cooling efficiency, and the insertion of renewable energy with the primary goals to reduce costs and improve sustainability. Additionally, enhanced energy management can result in reliability and security concerns. There are potential benefits to combining subsystem level efficiencies with total system sustainability and security solutions to develop holistic data center energy management policies.

9. ACKNOWLEDGEMENT

The authors would like to acknowledge Dr. Sara Zarei, Bill Kwai, and Mike Putnam for their support and involvement in literature review. This work was funded by SPAWAR Systems Center Pacific, Naval Innovation Science and Engineering (NISE) Program, Naval Research Enterprise Internship Program (NREIP), and the Office of Naval Research Energy Systems Technology Evaluation Program (ESTEP).

CITATIONS

[1] J. Koomey, Growth in Data center electricity use 2005 to 2010, Analytics Press (2011).
[2] Emerson Network Power, Maximizing Data Center Efficiency, Capacity and Availability through Integrated Infrastructure (2011).
[3] R. Bianchini, R. Rajamony, Power and energy management for server systems, Computer, Volume 37, Number 11 (2004):68-76.
[4] S. Greenberg, E. Mills, B. Tschudi, P. Rumsey, B. Myatt, Best practices for data centers: lessons learned from benchmarking 22 data centers, Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings in Asilomar, CA. ACEEE, August 3 (2006): 76-87.
[5] VMWare, VMWare Inc, Accessed Mar 2013, <http://www.vmware.com/products/vcenter-server/overview.html>.
[6] D. Kusnetzky, Virtualization: A Manager's Guide, O'Reilly Media Inc (2011), 2-3.
[7] P. Mell, T. Grance, The NIST Definition of Cloud Computing, NIST Special Publication, 800-145 (2011).
[8] M. Macias, J. Guitart, Client Classification Policies for SLA Enforcement in Shared Cloud Datacenters, *Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on* [2012] 156,163
[9] W. Huang, M. Allen-Ware, J. Carter, Tapo: Thermal-aware power optimization techniques for servers and data centers, Green Computing Conference and Workshops (IGCC), 2011 International. IEEE (2011).
[10] Y.H. Said, S. Bergeon, S. Ploix, Q.H. Giap, I. Dumitru, Towards a global energy management within a data center, in: IEEE 11th International Conference on Electrical Power Quality and Utilisation (2011).
[11] Q. Tang, S.K.,Gupta, G. Varsamopoulos, Energy-efficient thermal-aware task scheduling for homogeneous high-performance computing data centers: A cyber-physical approach. IEEE Transactions on Parallel and Distributed Systems, Volume 19, Number 11 (2008), 1458-1472.
[12] S. Pelley, D. Meisner, T.F. Wenisch, J.W. VanGilder, Understanding and abstracting total data center power, In Workshop on Energy-Efficient Design, 2009.
[13] Qualitative Analysis of Power Distribution Configurations for Data Centers, White Paper #4, The Green Grid, Accessed July 31, 2013, <www.thegreengrid.org>
[14] X. Fan, W.D. Weber, L.A. Barroso, Power provisioning for a warehouse-sized computer, ACM SIGARCH Computer Architecture News, Volume 35, Number 2 (2007), 13-23.
[15] TIA, TIA-942 Data Center Standards Overview, Telecommunications Industry Association (2005).
[16] N. Rasmussen, Electrical efficiency modeling of data centers, APC, White Paper #113 (2006), Accessed Aug 5, 2013, <media.comtec.com/pdf/apc/ElectricalEfficiencyModelling.pdf>
[17] D. Cavdar, F. Alagoz "A Survey of Research on Greening Data Centers," IEEE Global Communications Conference (2012).
[18] M. Stransberry, J. Kurdritzki. Uptime Institute 2012 Data Center Industry Survey, UptimeInstitute (2012).
[19] M. Iyengar, M. David, P. Parida, V. Kamath, B. Kochuparambil, D. Graybill, M. Schultz, M. Gaynes, R. Simons, R. Schmidt, T. Chainer, T., Server liquid cooling with chiller-less data center design to enable significant energy savings, Semiconductor Thermal Measurement and Management, Symposium (SEMI-THERM), 2012 28th Annual IEEE (2012) 212,223
[20] C. Tang, M. Steinder, M. Spreitzer, G. Pacifici, A scalable application placement controller for enterprise data centers, Proceedings of the 16th international conference on World Wide Web. ACM (2007).
[21] L. Liu, H. Wang, X. Liu, X. Jin, W.B. He, Q.B. Wang, Y. Chen, GreenCloud: a new architecture for green data center, Proceedings of the 6th international conference industry session on Autonomic computing and communications industry session. ACM (2009).
[22] I. Goiri, F. Julia, R. Nou, J.L. Berral, J. Guitart, J. Torres, Energy-Aware Scheduling in Virtualized Datacenters, Cluster Computing (CLUSTER), 2010 IEEE International Conference on (2010) 58-67.
[23] J. Berral, R. Gavalda, J. Torres, Adaptive Scheduling on Power-Aware Managed Data-Centers Using Machine Learning, Grid Computing (GRID), 2011 12th IEEE/ACM International Conference on , (2011) 66,73, 21-23.
[24] Xen Project, The Hypervisor, Accessed Aug 2013, <www.xenproject.org/developers/teams/hypervisor.html>.

[25] Intel DCM, Data Center Manager, Intel Inc., Accessed Mar 2013, <http://software.intel.com/sites/datacentermanager/>.

[26] iDRAC, Integrated Dell Remote Access Controller, Dell Inc., Accessed Apr 2013, <http://www.dell.com/Learn/us/en/555/solutions/integrated-dell-remote-access-controller-idrac?c=us&l=en&s=biz>.

[27] PowerEdge, Shop PowerEdge Dell Blade Servers, Dell Inc., Accessed Mar 2013, <http://www.dell.com/us/business/p/poweredge-blade-servers>.

[28] ACPI, ACPI: Advanced Configuration and Power Interface, ACPI, Accessed Dec 2012, <http://www.acpi.info/>.

[29] Installation Instructions- Class 3200 Meter Advanced KWH/Demand Meter State Alone and MMU, E-Mon, Accessed May 10 2013, <http://www.emon.com/products_class3200.html>.

[30] Hioki 3333 AC/DC Power HiTester, Hioki, Accessed Jul. 30 2013. <http://www.tequipment.net/Hioki3334.html?gclid=CPWHmN_HqrYCFad_Qgod_hAA3A>.

[31] APC, Metered Rack PDU, APC Inc., Accessed June10 2013, <https://www.apc.com/products/family/index.cfm?id=136>.

[32] Liebert DC Precision Cooling System, Emerson Network Power, Accessed Jul. 30 2013. <http://www.liebert.com/EN-US/PRODUCTS/PRECISIONCOOLING/LARGEROOMCOOLING/Pages/LiebertDSPrecisionCoolingSystem28-105kW.aspx>.

[33] Leibert, Emerson's Liebert Brand - Optimize Data Center Infrastructure, Leibert Inc., Accessed Jun 10 2013. <http://info.emersonnetworkpower.com/>.

[34] A. Alaraifi, Unearthing the Value of Sensor Information Systems for Managing Organizations IT Infrastructure, in: Sixth Asia Modelling Symposium (2012) 13,16, 29-31.

[35] A. Beloglazov, A. Jemal, and R. Buyya, Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing, Future Generation Computer Systems 28.5 (2012) 755-768.

[36] B. Aksanli, J. Venkatesh, T. Rosing, Using Datacenter Simulation to Evaluate Green Energy Integration, IEEE Computing Society (2012).

[37] R. Calheiros, R. Ranjan, A. Beloglazov, C. De Rose, R. Buyya, CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environment and Evlauation of Resource Provisioning Algorithms, CloudSim (2011).

[38] D. Kliazovich, P. Bouvry, Y. Audzevich, S. Khan, GreenCloud: A Packet-level Simulator of Energy-aware Cloud Computing Data Centers, in: Proceedings of IEEE Global Telecommunications Conference (2010).

[39] B. Aksanli, J. Venkatesh, T. Rosing, Datacenter Modeling and Simulation with Focus on Energy Efficiency and Green Energy Integration. IEEE Computer Special Issue on Modeling and Simulation of Smart and Green Computing Systems (2012).

[40] E. Pakbaznia, M. Pedram, Minimizing data center cooling and server power costs, Proceedings of the 14th ACM/IEEE international symposium on Low power electronics and design. ACM, (2009).

[41] R. Zhou, Z. Wang, C.E. Bash, A. McReynolds, C. Hoover, R. Shih, N. Kumari, R.K. Sharma, A holistic and optimal approach for data center cooling management, American Control Conference (ACC), 2011. IEEE (2011)

[42] Tileflow: Overview, Innovation Research, Inc., Accessed Jul. 30 2013, <http://inres.com/products/tileflow/overview.html>.

[43] CoolSim Homepage, CoolSim, Accessed Jul. 30 2013, < http://www.coolsimsoftware.com/>.

[44] Cisco, Data Center Power and Cooling, White Paper (2011).

[45] D. Meisner, B.T. Gold, T.F. Wenisch, PowerNap: eliminating server idle power, In Proc. of ASPLOS ’09 (2009) 205–216.

[46] C. Lefurgy, X. Wang, M. Ware, Server-Level Power Control, Autonomic Computing, 2007. ICAC '07. Fourth International Conference on , vol., no., (2007) 4,4, 11-15 June 2007.

[47] F. Palmieri, S. Ricciardi, U. Fiore, Evaluating Network-Based DoS Attacks under the Energy Consumption Perspective: New Security Issues in the Coming Green ICT Area, International Conference on Broadband and Wireless Computing, Communication and Applications (2011) 374-379.

[48] A. Beloglazov, Energy-Efficient Management of Virtual Machines in Data Centers for Cloud Computing, Doctoral Dissertation, University of Melbourne, 2013.

[49] A. Banerjee, T. Mukherjee, G. Varsampoulos, S.K.S. Gupta, Cooling-aware and thermal-aware workload placement for green HPC data centers, Green Computing Conference. (2010)

[50] S. Harizopoulos, M. Shah, J. Meza, P. Ranganathan, Energy efficiency: The new holy grail of data management systems research. arXiv preprint arXiv:0909.1784 (2009).

[51] C. Li, A. Qouneh, T. Li, iSwitch: coordinating and optimizing renewable energy powered server clusters, Computer Architecture (ISCA), 2012 39th Annual International Symposium on, IEEE, 2012.

[52] C. Li, W. Zhang, C.B. Cho, T, Li, SolarCore: Solar energy driven multi-core architecture power management. In: High Performance Computer Architecture (HPCA), 2011 IEEE 17th International Symposium on, pp.205-216.

[53] P. Bohrer, E. Elnozahy, T. Keller, M. Kistler, C. Lefurgy, and R. Rajamony, The case for power management in web servers, Power Aware Computing (2002).

[54] I, Goiri, F. Julia, J. Guitart, Efficient data management support for virtualized service providers, In: Parallel, Distributed and Network-based Processing, 2009 17th Euromicro International Conference on, IEEE, 2009.

[55] “Omnet,” 2009, http://www.omnet.org.

[56] J.L. Berral, I. Goiri, R. Nou, F. Julia, J. Guitart, R. Gavalda, J. Torres, Towards energy-aware scheduling in data centers using machine learning. Proceedings of the 1st International Conference on energy-Efficient Computing and Networking, pp. 215-224, (2010), ACM.

[57] M. Green, S. Karajgikar, P. Vozza, N. Gmitter, D. Dyer, Achieving energy efficient data centers using cooling path management coupled with ASHRAE standards, Semiconductor Thermal Measurement and Management Symposium (SEMI-THERM), 2012 28th Annual IEEE, (2012) 288,292

[58] J. Rambo, J. Yogendra, Modeling of data center airflow and heat transfer: State of the art and future trends, Springer Science (2007) .

[59] R. Schmidt, E. Cruz, E.E, M. Iyengar, Challenges of data center thermal management, IBM Journal of Research and Development , vol.49, no.4.5, (2005) 709,723

[60] R.R. Schmidt, Thermal Profile of a High Density Data Center – Methodology to Thermally Characterize a Data Center, NA-04-4-2, ASHRAE Transactions, Vol. 110, Iss. 2, (2004) 635.

[61] D.C. Hwang, V.P. Manno, M. Hodes, G.J. Chan, Energy savings achievable through liquid cooling: A rack level case study, Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2010 12th IEEE Intersociety Conference (2010)

[62] Google Data Centers: Renewable Energy http://www.google.com/about/datacenters/renewable/ Accessed: July 9, 2013

[63] Apple and the Environment: Powering our facilities with clean, renewable energy http://www.apple.com/environment/renewable-energy/ Accessed: July 9, 2013

[64] Geothermal Energy, Natural Resources Defense Council, Accessed Aug 1, 2013, <www.nrdc.org/energy/renewables/geothermal.asp#footnote2>

[65] Variability of Renewable Energy Sources, National Renewable Energy Laboratory, Accessed Aug 1, 2013, <www.nrel.gov/electricity/transmission/variability.html>

[66] C. Stewart, K. Shen, Some joules are more precious than others: Managing renewable energy in the datacenter, in:

Proceedings of the Workshop on Power Aware Computing and Systems, (2009)
[67] N. Deng, C. Stewart, J. Li, Concentrating renewable energy in grid-tied datacenters, in: IEEE International Symposium on Sustainable Systems and Technology, (2011).
[68] Í. Goiri, R. Beauchea, K. Lee, T.D. Nguyen, GreenSlot: scheduling energy consumption in green datacenters, in: Proceedings of International Conference for High Performance Computing, Networking, Storage and Analysis, (2011).
[69] K. Le, R. Bianchini, M. Martonosi, T. Nguyen, Cost-and-energy-aware load distribution across data centers. Proceedings of HotPower (2009), pp.1-5.
[70] Y. Gao, Z. Zeng, X. Liu, P.R. Kumar, The Answer Is Blowing in the Wind: Analysis of Powering Internet Data Centers with Wind Energy, in: Proceedings of INFOCOM (Mini Conference), (2013).
[71] Z. Liu, M. Lin, A. Wierman, S.H. Low, L. Andrew, Greening geographical load balancing, in: Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems, (2011).
[72] Y. Zhang, Y. Wang, X. Wang, GreenWare: greening cloud-scale data centers to maximize the use of renewable energy, Springer Berlin Heidelberg Middleware (2011), 143-164.
[73] M. Lin, Z. Liu, A. Wierman, L.L.H. Andrew, Online algorithms for geographical load balancing, in: IEEE International Green Computing Conference (IGCC), (2012).
[74] C. Chen, B. He, X. Tang, Green-aware workload scheduling in geographically distributed data centers, in: IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom), (2012).
[75] M. Ghamkhari, H. Mohsenian-Rad, Optimal integration of renewable energy resources in data centers with behind-the-meter renewable generator, in: IEEE International Conference on Communications (ICC), (2012).
[76] S. Akoush, R. Sohan, A. Rice, A.W. Moore, A. Hopper, Free lunch: exploiting renewable energy for computing, HotOS (2011).
[77] Z. Liu, A. Wierman, Y. Chen, B. Razon, Ni. Chen, Data center demand response: Avoiding the coincident peak via workload shifting and local generation, in: Proceedings of the ACM SIGMETRICS/international conference on Measurement and modeling of computer systems, (2013).
[78] D. Terdiman "Stuxnet delivered to Iranian nuclear plant on thumb drive." http://news.cnet.com/8301-13772_3-57413329-52/stuxnet-delivered-to-iranian-nuclear-plant-on-thumb-drive/
[79] D. Seo, H. Lee, A. Perrig, Secure and Efficient Capability-Based Power Management in the Smart Grid, Parallel and Distributed Processing with Applications Workshops (ISPAW), 2011 Ninth IEEE International Symposium (2011) 119-126
[80] R.B. Vaughn, R. Henning, A. Siraj, Information Assurance Measures and Metrics-State of Practice and Proposed Taxonomy, Proceedings of the 36th Hawaii International Conference on System Sciences (2002).
[81] H. Tram, Technical and operation considerations in using Smart Metering for outage management, Transmission and Distribution Conference and Exposition, 2008. T&D. IEEE/PES (2008) 1-3, 21-24
[82] K. Kant, M. Le, and S. Jajodia, Security considerations in data center configuration management, Configuration Analytics and Automation (SAFECONFIG), 2011 4th Symposium on. IEEE (2011).
[83] H. Hlavacs, T. Treutner, J.-P. Gelas, L. Lefevre, A. C. Orgerie, Energy Consumption Side-Channel Attack at Virtual Machines in a Cloud, Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference (2011) 605-612.

APPENDIX B: A Computing Testbed

Draft to be submitted to 2014 IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing (CCGrid) Paper Due: 11 Nov 2013

Computing energy testbed:

A model for the evaluation of energy management techniques for data centers

Adriane Wolfe, James Jen, Ryan Huynh
Space and Naval Warfare Systems Center-Pacific
53560 Hull Street, San Diego CA, 92152-5001
adriane.wotawa-berge@navy.mil
james.jen@navy.mil

Daniel Lyons
Department of Mathematics and Statistics
San Diego State University
5500 Campanile Dr. San Diego, CA 92182-7720
danielclyons@gmail.com

Abstract— **Recent years have seen a large body of theoretical research on techniques for energy efficiency in data centers equipped with virtualization. However, little work has focused on obtaining experimental results from an actual computing testbed to demonstrate true energy savings. The purpose of the present work is twofold: First, this paper describes the design and present capabilities of a computing energy testbed. Second, this work describes experimental results for VM loading experimentation obtained from the testbed. The results confirm that the computing energy testbed is a valuable resource for testing energy efficiency techniques.**

Keywords— Data center; energy; testbed; virtualization.

INTRODUCTION

Energy consumption by data centers has surged in recent times: from 2005 to 2010, the energy used by data centers increased by 36%, and in 2011, data centers accounted for approximately 2% of total electricity use in the U.S. [1:Koomey, 2011]. This increase in energy consumption has sparked a movement that seeks to develop techniques for improving the energy efficiency of data centers. The computing energy testbed (CET) has been developed in an effort to develop and apply energy efficiency techniques for rapid deployment in high performance virtual data centers. The goal is to both test efficiency as well as ensure reliable and secure performance so that approaches can be transitioned into Navy operational systems.

This paper is organized as follows: Section II discusses the relevant background information that pertains to data center energy issues. Section III contains the approach for the development of the CET and outlines the initial experiment performed to evaluate the CET. Section IV presents the results obtained from the outlined experiment. Section V draws conclusions about the research, and finally, section VI discusses our future work.

BACKGROUND

A data center is a facility that provides computing resources on demand as required. The most fundamental systems in a data center are computing (IT), cooling, power distribution, and management systems. In order to achieve energy efficiency in a data center, it is often helpful to consider energy saving tactics specific to the systems.

Information Technology (IT)

The IT equipment is the most fundamental component of the data center, as it directly provides computing services to the client. Approximately 70% of IT power consumption is due to the server, P_{Server}, which accounts for power consumption due to CPU, memory, disks, peripherals, and the motherboard [Cavdar, 2012]. Other important components of IT power consumption are data storage, P_{SAN}, and power for the network equipment for data communication, $P_{Network}$ [Cisco, 2011], so that the total IT power obeys the equation

$$P_{IT} = P_{server} + P_{SAN} + P_{Network} \quad \text{(Eq. 1)}$$

A large portion (at least 20%) of server power is consumed by the CPU [Meisner, 2009]. It is generally accepted that the CPU power consumed, P_{CPU}, is a linear function of CPU utilization, u, [Beloglazov, 2012]:

$$P_{CPU}(u) = k * P_{max} + (1 - k) * P_{max} * u \quad \text{(Eq. 2)}$$

where P_{max} is the maximum power consumed by the CPU when it reaches 100% utilization, k is the fraction of power consumed by the idle server (at 0% utilization), and u is CPU utilization [Beloglazov, 2012]. The power consumed by the CPU when idle is $k * P_{max}$, which may be approximately 60% of P_{Max} [Lefurgy, 2007].

Common techniques for improving energy efficiency include server mode scheduling, VM dynamic scheduling, dynamic voltage and frequency scaling, and power capping. Our previous work features an in-depth review of energy efficiency for IT systems [Wolfe, 2013, submitted].

This work is supported by SPAWAR Systems Center Pacific, Naval Research Enterprise Internship Program (NREIP), and Naval Innovation Science and Engineering (NISE) Program.

Note that these techniques are typically applied in order to obtain energy efficiency while satisfying the data center service level agreement (SLA), a specific contract outlining the computational obligations of the data center to supply its client(s) with computing service.

Cooling

Another significant source of data center power consumption is due to cooling technology installed in the data center to regulate the humidity and temperature appropriately for IT equipment operation. There are many technologies made available for cooling by the heating, ventilation, air conditioning (HVAC) industry. Most cooling systems consist of at least one of the following fundamental elements: heat transport, heat exchange, compressor and evaporative assist devices. Heat transport devices and compressors cool the data center by consuming electrical energy, while the evaporative assist requires steady water input. A cooling device that is installed in a data center is commonly known as a computer room air conditioner (CRAC) unit, which monitors and maintains the temperature, air distribution, and humidity in a data center by transport of chilled air [58: Rambo, 2007].

In order to maintain proper data center temperature, cool air is circulated throughout the data center in order to remove heat generated by the computing equipment [59: Schmidt, 2005]. This process is carried out by the CRAC as follows: cool air is initially pumped directly into the room or through an elevated floor with perforated tiles into a space between server racks known as the “cold aisle”. The heat generated by the computing equipment is absorbed by the incoming cool air and escapes into the “hot aisle”, where it is transferred to a cooling liquid [60: Schmidt, 2004]. The hot air is then either transported out of the room or supplied to a chiller unit which cools the air to prepare it for recirculation.

Some data centers utilize “free cooling” from outside air or liquid cooling for high thermal transport [61: Hwang, 2010]. While management of such cooling systems is technology-specific, the management of cooling is universally required.

The fundamental techniques that are used to improve energy efficiency in data center cooling systems are regulating supplied air temperature, adjusting the CRAC frequency, directed cool air delivery, and computational task scheduling. A review of data center relevant research on cooling techniques is contained in our recent paper [Wolfe, 2013, submitted].

Power Distribution

Power is generally provided via the electrical grid and backup via short term storage and mid-term energy generation. The short term storage is generally provided by uninterruptable power supplies (UPS) which are large batteries with voltage regulation that often provide backup power on the scale of minutes to hours. Diesel fuel backup generators with fuel tanks are used to generate power on the scale of hours to days. In addition to backup energy sources data centers often have redundant power architectures, where the power distribution, storage, and generation components and feeds are redundant to reduce the impact of device failures and to provide alternative power supply during maintenance.

Power distribution is generally comprised of transformers which step down grid power to building/data center level voltages (generally 3 phase 480 VAC and 3 phase 208 VAC). CRACS and chillers run off of these voltage levels, while servers run off of either 208 VAC or 120 V single phase. Data centers often use 208 VAC over 120 VAC because it is more energy efficient and can distribute higher power with the current breaker levels. IT equipment hosted in racks are powered by either rack power distribution units (PDU) or by ties to busways.

APPROACH

Testbed

The Computing Energy Testbed (CET) was established in order to support rapid technology adoption for data center energy efficiency and security. The CET enables rapid technology adoption by providing a flexible computing environment where technologies and approaches can be developed, evaluated, and validated prior to operational testing. It is modeled as a “closet” scale data center and consists of a small contained computing room with a large connected room for systems operations management and monitoring Fig. 1.

The computing environment consists of two populated racks and two unpopulated racks as seen in Fig. 2, Fig. 3. The CET computing environment is fully virtualized using VMWare. One lab host server hosts lab software including data center infrastructure management (DCIM) and network security software. Two server racks are populated with servers, storage area network (SAN), and rack mounted uninterruptable power supplies (UPS). An additional two racks are currently empty but are available for future growth, and provide row architecture for cooling. Rack 1, is used for full experimentation and can be configured to operate virtual machines running bench mark applications. Rack 1 hosts five dell power edge 2850s one dell power edge 2950. In contrast, Rack 2 equipment is used for a system baseline and is configured by application owners for specific Navy operational requirements. Rack 2 hosts five HP Proliant DL58037 servers.

Figure 1: Lab systems operations management and monitoring.

Figure 2: CET physical computing environment.

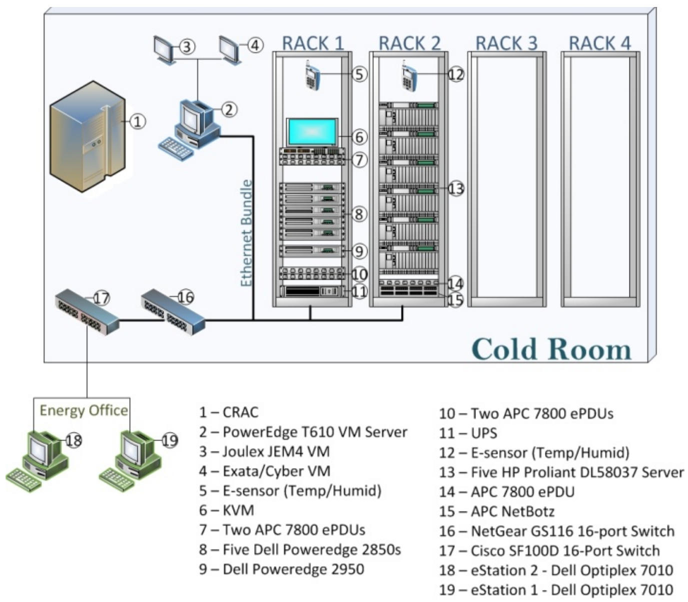


Figure 3: CET physical and network architecture.

A large body of research exists that recommends the implementation of all possible energy efficiency techniques. However, from the data center perspective, it is not always possible to implement every energy management technique, since the data center typically does not have complete control of all datacenter functionality. For example, data centers that provide clients with computing resources via the Software as a Service (SaaS) paradigm have direct control of data center servers, so they can implement all energy efficiency techniques, i.e., virtual machine allocation, putting servers in sleep mode, dynamic voltage and frequency scaling. In contrast to SaaS, data centers that provide resources via the Infrastructure as a Service (IaaS) paradigm rent out their equipment to the client, so that the client retains control of the computing resources, and very few energy efficiency techniques may be implemented by the data center. The multi-rack layout of the CET provides the infrastructure to test the various service level paradigms.

The power distribution architecture is scaled down from an operational data center (Fig. 4). The racks are powered by metered rack power distribution units (PDUs). Rack 1 PDUs (APC Metered PDU, APC Switched PDU) are powered by rack mounted UPS (APC, Smart UPS), which are powered by the 120 V single phase power distribution. Rack 2 equipment is currently powered by 120 V via the rack PDU but is scheduled to shift to 208 V in the near future. A computer room air conditioner (CRAC) unit (Liebert Challenger 3000) is powered by 208 V and provides cool air to the room which is located in the room to blow air to the intake (cool aisle) side of the racks. The CRAC and power to the racks are metered by power smart meters (E-Mon D-Mon).

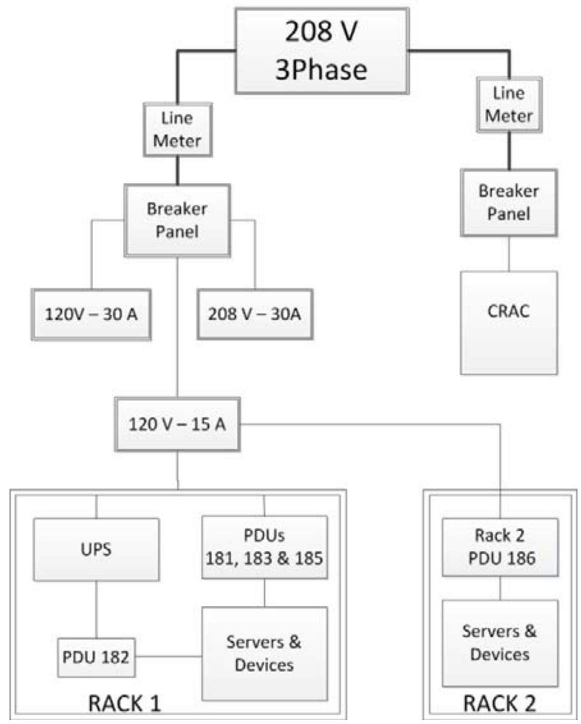


Figure 4: CET power distribution architecture.

The operations and monitoring room provides monitors that display JEM (JouleX Energy Manager), a DCIM dashboard that can be used to monitor and configure the smart metering (Figure D). The smart metering includes monitoring the rack PDUs, servers, VMWare, environmental sensors, CRAC, and power meters. Networked environmental meters include APC Temperature and Humidity Sensors with and without display connected to APC Netboltz Rack Monitor. Additional metering is conducted via iDRAC, power quality analyzers (Hioki 3333, Fluke 434-II Power Quality and Energy Analyzer), power meters (E-Mon D-Mon Class 3200 Smart Meter), and a thermal imager (Fluke TI125). The thermal environment and conditions can be modeled using TileFlow (TileFlow, Innovative Research Inc.) and the network security of the smart metering network is being modeled using Xcada.

Computing Efficiency

A computing efficiency experimental setup was created comparing computing utilization to power consumption.

▪ *Experimental Setup*

The VMWare ESXi hypervisor were installed on the Dell PowerEdge 2950 and eight virtual machines were created. Each virtual machine hosted a Red Hat Linux 6.3 operating system with an allocation of 4GB of RAM and 40GB of hard drive space. Each VM hosted on application of STRESS, a computing benchmark application that is deliberately a simple command-line program and works by numerically intensive square root command to compute the square root of a random number. To test the validity of the computing energy testbed, the idle Dell Power Edge 2950 server is stepped up from 0 to 8 virtual machine (VM) levels, spending approximately five minutes at each level.

Figure 5: Dell power edge 2950 server used for experimentation.

▪ *Experimental Data Collection*

The goal was to gather power and utilization data for the server from Joulex Energy Manager (JEM). This would allow us to match up CPU utilization with its power expenditure with correct time stamping. While the server is configured on the JouleX dashboard, it was found that it could sample data at 1 minute intervals, which was not granular enough for dynamic monitoring of VM loading. In addition JouleX, in many cases, only estimates the power usage of a server. Instead the server power consumption data was directly measured with a Hioki Model No. 3334 Power Analyzer (Fig. 7. The power analyzer measured in real time and high sampling rate, the energy expenditure of the server and logs it into a file on the computer.

The average power input to the server from the power distribution unit (PDU) is measured using the Hioki, 3334 with a 2.000 KHz sampling rate. The CPU usage (utilization), both by core and total, is measured (or estimated) by VXphere Client for all CPU usage and separately each of the individual CPU's core usage at a 0.05 Hz sampling rate.

Figure 7: Hioki 3334 Power Meter

RESULTS

Experimental Results

The average power input to the server is plotted in Fig. 8. A low-pass filter was used to smooth the power data so that clear steps between VM levels were observed (not shown). The power data is down-sampled to 0.05 Hz so that the same sampling rate is used for comparison with the CPU usage data (Fig. 9).

Because the data was collected by two sources, aligning the two data types with respect to time was challenging. The two data sets were aligned by the following step. A low pass filter was applied to the power data so that each VM level transition, indicated by large slope, is readily apparent (Fig. 8). The time derivative of the filtered power data was taken, which graphically revealed the 8 locations of VM level changes by local maxima of large derivative (threshold value: above 0.4 W/sec). Step 2 was repeated for the usage data with threshold 0.15 %/sec (filtering not necessary due to obvious VM level changes and low sampling rate (0.05 Hz)). In order to obtain a direct comparison between power and usage data, the power data was downsampled by a factor of 200-since 10 Hz/0.05 Hz = 200- to establish a one-to-one correspondence between power and usage data. The correlation coefficient of the usage data (0.05 Hz) and the down sampled power data (at 0.05 Hz) was calculated with the corrcoef() function in Matlab to determine how closely the power and usage curves match. The corrcoef() function in Matlab outputs a matrix that is calculated from the covariance matrix consisting of the power and usage data [Lee Rodgers]. The usage data was shifted backward one time step (20 sec) and Step 5 was repeated. This procedure was done repeatedly, with the usage data shifted backward one time step each time. The alignment of the usage and power data (at 0.05 Hz) with the maximum correlation coefficient (closest to 1 in absolute value) was determined to be the most correct alignment possible with the given data.

However, it is acknowledged that some error can be caused based on the uncharacterized transient response of the CPU to power performance relationship.

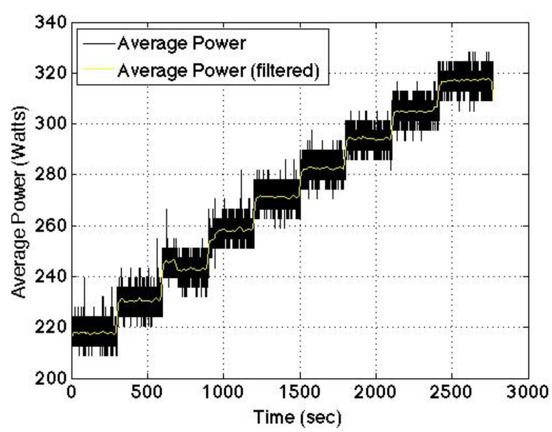


Figure 8: Hioki 3334 Power Meter raw power data

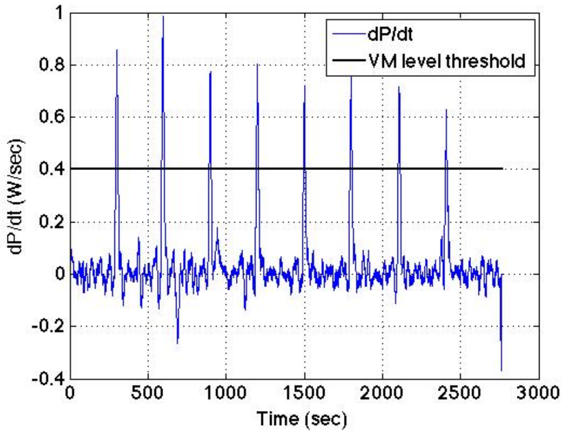


Figure 9: Derivative of power data (dP/dt)

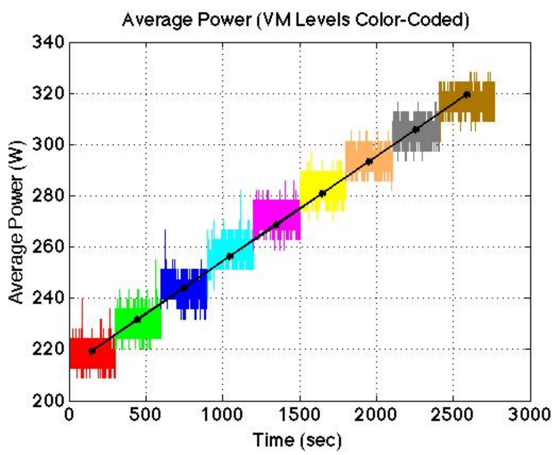


Figure 10: Average Power Data color-coded by VM Level.

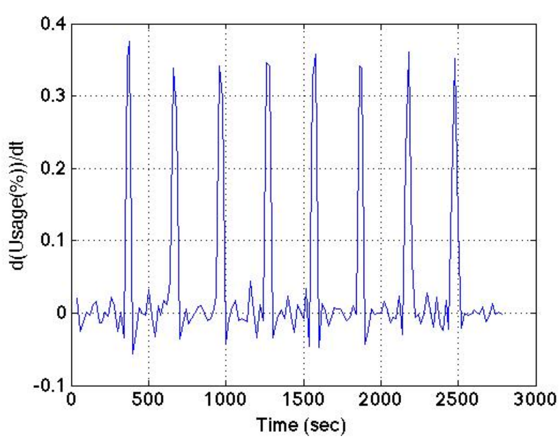


Figure 11: Derivative of utilization data (du/dt)

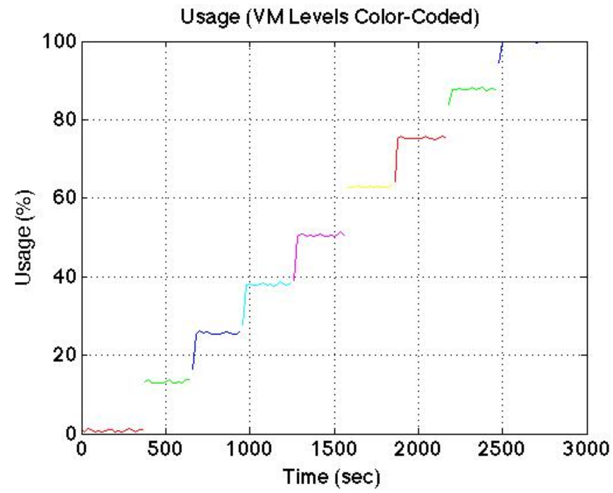


Figure 12: CPU utilization data color-coded by VM Level.

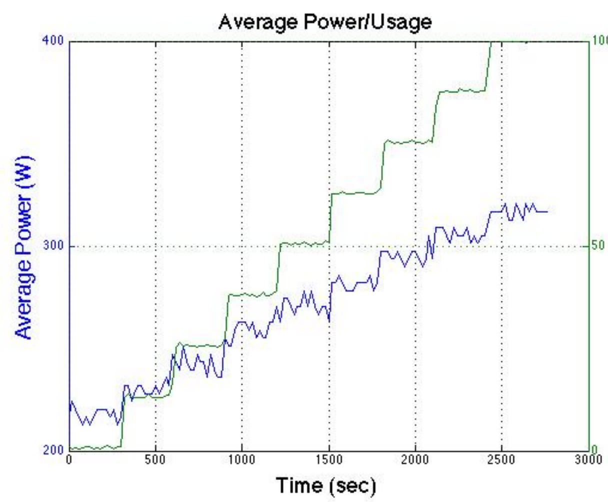


Figure 13: Power correlated to CPU utilization data.

In order to verify the linear relationship between server power and CPU usage, the mean power and mean usage are calculated at each VM level. A linear least squares fit of CPU usage verses power is performed in MATLAB, and the constants $P_{max} = 318.99\,W$ and $k = 0.68499$ are obtained for the Equation 2. The correlation coefficient of the observed mean power data and the predicted mean power data obtained from this equation is 0.99899, which indicates very good agreement between the actual data and the linear equation above.

One metric for energy efficiency for power servers is the ratio of power and CPU usage. Solving for P/u in the previous linear equation, we obtain Eq. 3,

$$\frac{P}{u} = \frac{b}{u} + a \qquad \text{Equation 3}$$

where $b = kP_{max}$ and $a = (1-k)P_{max}/100$ where the constant a has been converted from a percentage to a decimal. The parameters k and P_{max} are identical to those used in the above linear fit. Thus, for the current data set, the parameter values $a = 1.0049\ W/\%$ and $b = 218.5043\ W$ are obtained. A comparison of the actual data with the curve fit is made in Fig. H. Each data point in the plot is in Watts per CPU usage percent, so that the plot essentially represents a cost function. A straightforward observation from the plots is that the power required to increase CPU usage by exactly one percent decreases as more VMs are initialized. It is reasonable to conclude that the energy efficiency of the server increases as the number of VM initialized increases.

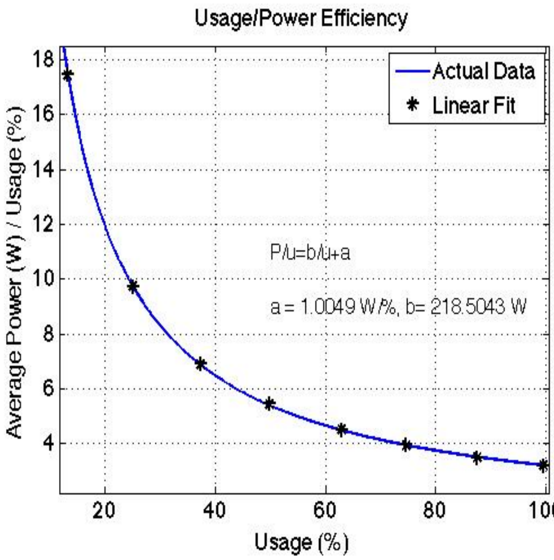


Figure 14: A zoomed-in view of the top plot (without the 0 VM point).

The CPU usage-obtained by JEM-is plotted in the figure below as a percent of total server usage by core. The plot below shows that the physical cores share the computing load using an internal energy/performance management strategy.

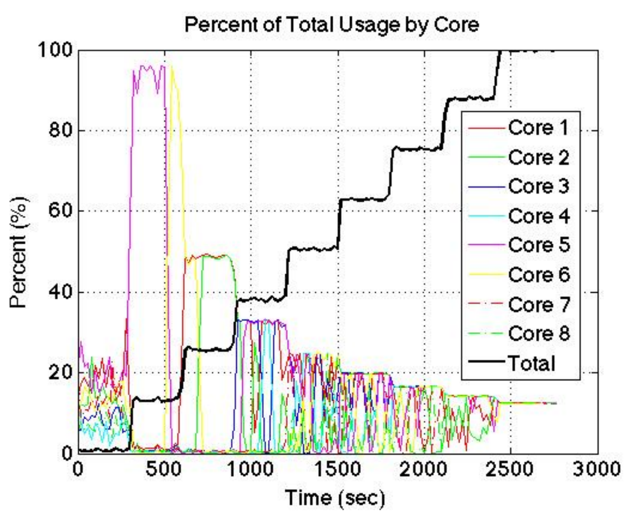


Figure 15: (Black) The percentage of total CPU usage of all 8 cores is plotted. The eight remaining curves give the percent usage by core of all cores.

Testbed

The testbed is evaluated in terms of the efficiency and management techniques it can support. The computing efficiency experiment demonstrated the ability to configure and manage VM status, collect server power and CPU utilization. The data collected correlated strongly with modeled expected performance and was on trend with regard to similar literature findings. The disparate data collection devices had inconsistent time-delays, however using the process described, data was correlated, and results where closely aligned with performance model.

CONCLUSION

The computing energy testbed (CET) was developed in response to the need for rapid development of energy efficiency techniques in data centers. The CET is comprised of IT, cooling, power distribution infrastructure. The energy consumption of the IT infrastructure is monitored by the JouleX Energy Manager (JEM), while the smart metering technology assists in monitoring power consumption of all data center components. Our virtual machine dynamic allocation experiment demonstrated the capability of the CET to simulate a virtualized datacenter at small scale.

FUTURE WORK

An important future task is conducting live migrations between servers, as presently, most live migration research is simulation-based. In order to perform such research experimentally, we must first install the developer version of VMWare() [citation]. Additionally, data collected from different sources has potential timing errors. The DCIM solution must be managed to collect accurate, synchronized data at granular time intervals.

Another focus of our upcoming research is the utilization of our overall CET infrastructure to design and evaluate datacenter energy efficiency techniques. The smart metering infrastructure provides the opportunity to experimentally evaluate energy efficiency techniques for IT, cooling, and power distribution, as the power utilization effectiveness (PUE) may be directly calculated for the CET. The CRAC system must be networked and controlled by the DCIM software and a feedback control loop must be designed between the CRAC and temperature sensors to support optimization of the CRAC temperature setpoint. To reduce mixing, a cold aisle containment should be added. With these additions holistic cooling and VM allocation testing can be conducted.

A forthcoming work will implement energy efficiency strategies in the CET and compare and contrast directly calculated data center energy savings that can be obtained at the different datacenter service levels without violating the data center service level agreement.

REFERENCES

[1] [Koomey, 2011] J. Koomey, Growth in data center electricity use 2005 to 2010, Analytics Press (2011).

[2] [Cavdar, 2012] D. Cavdar, F. Alagoz "A survey of research on greening data centers," IEEE Global Communications Conference (2012).

[3] [Cisco, 2011] Cisco, Data center power and cooling, White Paper (2011).

[4] [Beloglazov, 2012] A. Beloglazov, A. Jemal, and R. Buyya, Energy-aware resource allocation heuristics for efficient management of data

centers for cloud computing, Future Generation Computer Systems 28.5 (2012) 755-768.

[5] [Lefurgy, 2007] C. Lefurgy, X. Wang, M. Ware, Server-Level Power Control, Autonomic Computing, 2007. ICAC '07. Fourth International Conference on , vol., no., (2007) 4,4, 11-15 June 2007.

[6] [Wolfe, 2013] A. Wolfe, J. Jen, D. Lyons, C. Chen, V. Dinh, J. Romero-Mariona, C. In, S. Curtis, “Review of scheduling-based data center energy management,” unpublished.

[7] [Rambo, 2007] J. Rambo, J. Yogendra, Modeling of data center airflow and heat transfer: State of the art and future trends, Springer Science (2007).

[8] [Schmidt, 2005] R. Schmidt, E. Cruz, E.E, M. Iyengar, Challenges of data center thermal management, IBM Journal of Research and Development , vol.49, no.4.5, (2005) 709,723.

[9] [Schmidt, 2004] R.R. Schmidt, Thermal Profile of a High Density Data Center – Methodology to Thermally Characterize a Data Center, NA-04-4-2, ASHRAE Transactions, Vol. 110, Iss. 2, (2004) 635.

[10] [Hwang, 2010] D.C. Hwang, V.P. Manno, M. Hodes, G.J. Chan, Energy savings achievable through liquid cooling: A rack level case study, Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2010 12th IEEE Intersociety Conference (2010).

[11] Lee Rodgers, J., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. The American Statistician, 42(1), 59-66.

APPENDIX C: Siemens High School Paper

Hybrid Photovoltaic Energy Management for Data Center Application: A look toward cost competitive solutions

High School Student: Bill Kwai, **Mentor:** Adriane Wolfe

1: INTRODUCTION

Data centers are dedicated areas where organizations house information technology (IT) systems to host computing applications. Due to associated computing and cooling loads, data centers can consume up to one hundred times the energy of an office building of the same size [1]. In 2007, the EPA estimated that the national energy consumption from data centers in 2006 was 61 billion KWh, or 1.5 percent of the nation's total energy consumption, which is the equivalent of 65 million average American households for a month [2, 3]. This demand has grown, and in 2010, it was estimated to be 2 percent [4]. This energy consumption has associate operational costs and environmental impact, which has motivated research to reduce data center grid energy consumption.

Large corporations, such as Google, Apple, and Facebook have been paving the road for reduced brown, or grid, energy consumption via energy efficiency and renewable energy integration [5, 6, 7, 8]. Their efforts are mainly targeted at their large scale data centers. Eighty-two percent of data center associates say that their efforts toward energy usage reduction is based on reducing costs [9]. The goal of this paper is to explore a technique that allows medium and small sized data centers to reduce their brown energy consumption via Photovoltaic integration. A data center that is outfitted with a solar array on the rooftop was modeled. It is hypothesized that by effectively managing the solar generated power, via battery storage, end users will see reductions in energy costs resulting in a 10 year return on investment (ROI). By demonstrating approaches where photovoltaic integration in data centers is financially beneficial, data centers of small to medium scale can consider renewable energy technology adoption, thereby reducing both brown energy consumption and greenhouse gas emissions.

2: METHODS

2.1: Common Procedures

The traditional approach to adoption of photovoltaic (PV) systems is to grid tie the PV systems to the load (i.e. household or facility), which directly uses the energy generated. Any excess electricity that is generated is sold back to the utilities company [10]. When the excess energy is sold back through the grid, the end users are reimbursed at a wholesale rate, which is much less than what utilities customers pay for electricity. Only some companies offer net metering, which allows customers to sell excess solar energy at a retail price [10]. This is simple, and does not involve an expensive storage device, but it does not leverage utilities pricing agreements in order to maximize the operational cost reduction.

To address the concerns regarding the potential of having either excessive generation, or reduced generation due to variability of PV, there are two popular and widely researched methods of managing systems for renewable energy integration: geographical load balancing, and demand response techniques [11, 12, 13, 14, 15, 16]. In geographical load balancing, computing services are carried out at different geographical locations depending on the availability of energy. Demand response techniques encompass a broader array of methods to downshift energy consumption during periods of time when energy is more expensive or renewable energy is not available. These techniques can be used to optimize operational cost reduction. However, they are not always practical to employ.

2.2: Procedure

There were four main steps in the procedure developed, (1) evaluation of the Energy System Model, (2) characterization of data center energy load, (3) refinement of energy storage, and (4) hybrid management approach based on two pricing models. The two management approaches developed were specifically tailored for Time of Use and Coincident Peak agreements.

Energy System Model

A 5,000 square foot model data center outfitted with a Photovoltaic system on the roof was used [17]. The system stored energy in a battery bank and it was assumed that the data center load could draw power from both the grid and the battery via a power combiner, which handled synchronization, with no backflow into the grid [18].

Data was collected from a variety of published and government sources. Solar energy generation data was based off of NREL's PVWatts web application [19]. PVWatts uses historical irradiance values in select geographical locations in each state, and provides users with an hourly solar energy generation value for everyday of every month. In addition, data sets obtained from PVWatts were de-rated by 68.6% to account for the hybrid-energy system losses including those associated with inverters, transformers, and battery storage [20, 21]. The de-rating factor presents a realistic scenario of the usable amount of solar energy within the data center.

The photovoltaic array was modeled with a Commercial off the Shelf (CotS) product - the Suniva MVX250-60-5-701 solar module [22]. Rated at 250W DC, each panel stands at 64.54inches x 39.05inchees in length and width, so each module takes up approximately 17.50 square foot. Each module is then partitioned 25 square foot on the roof to account for walkways and shading from consecutive modules. Using this data, the 5,000 square foot rooftop allows for 200 solar modules. This design provides a 50kW system.

Next, battery storage devices were explored, and four types of batteries were studied: flooded lead-acid batteries, absorbed gas mat(AGM) lead-acid batteries, sealed lead-acid batteries, and lithium-ion batteries [23]. Among lead-acid batteries, flooded lead acid batteries were selected because they are the least expensive among sealed and AGM batteries, even though they require occasional maintenance [24].

Finally, the data center was modeled on a one line power diagram using the ETAP software, shown in Figure 1.

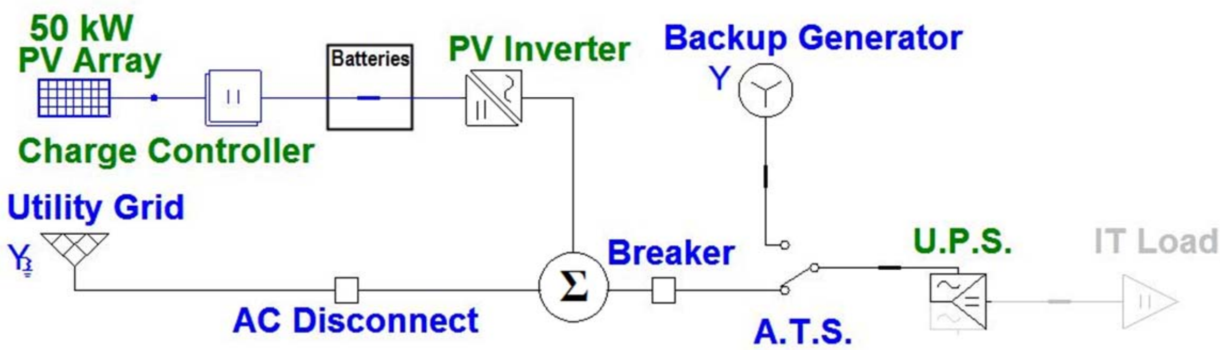


Figure 1: Model Data Center One Line Power Diagram

Data Center Energy Load Model

In order to create a realistic model, all energy management techniques are applied to the load curve of a NetApp data center building (Fig. 2). The load curve of the NetApp data center was metered by a study from the Lawrence Berkeley National Laboratory, so the curve represents the power consumption of a real data center [15]. The data set values from NetApp are then scaled to fit the data center power consumption of the 5,000 square foot data center using Equation 1:

$$\overline{E}_{netapp} \times k = \overline{E}_{model} \quad \text{(Equation 1)}$$

where E_{netapp} is the average energy consumption for the Netapp data center, k is the scaling constant, and E_{model} is the average energy consumption for the model data center.

In the model data center, it is assumed that the facility is outfitted with 210 server racks [17]. After scaling, the data points are fitted to a sixth degree polynomial (Fig. 3).

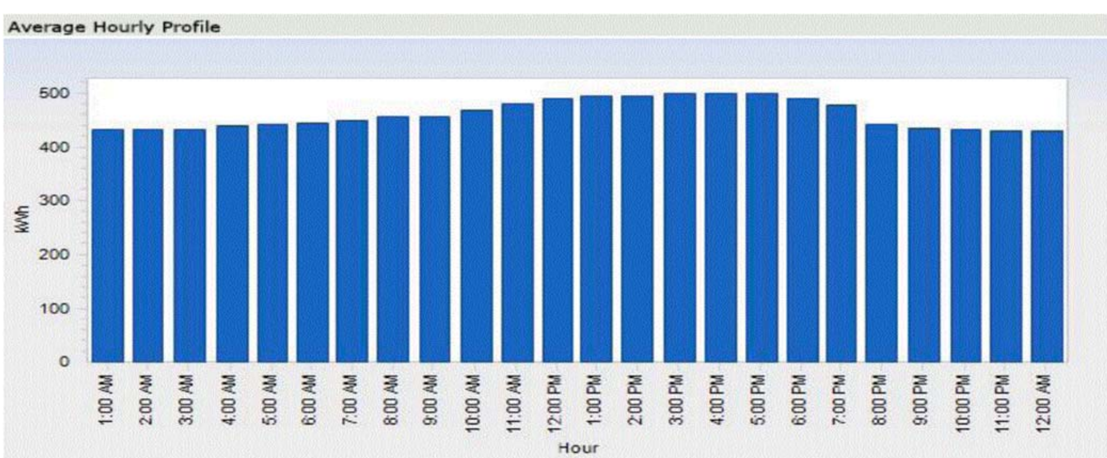


Figure 2: NetApp data center load curve [15]

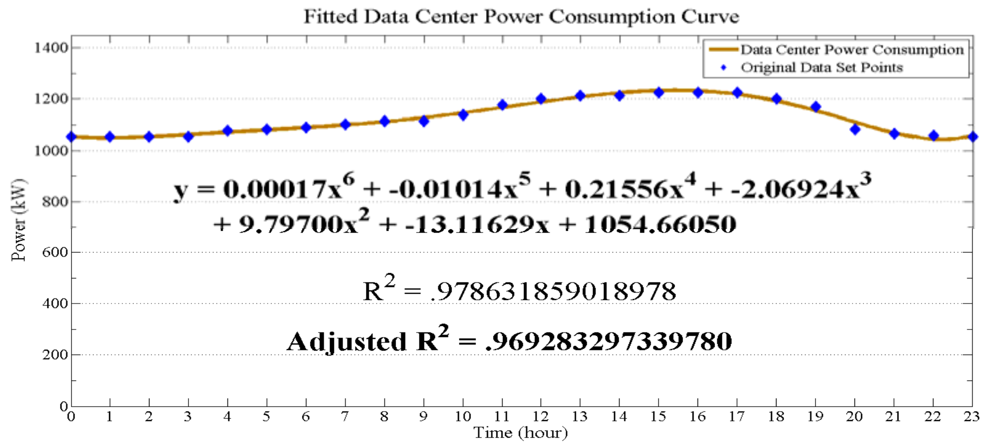


Figure 3: Fitted NetApp data center load curve

For all techniques and analysis, it is assumed that the polynomial fitted curve is representative of the actual data center load. In addition, it is assumed that the load of the data center did not vary throughout different seasons. While the assumption that the IT load remains relatively constant seems to be a reasonable, it is acknowledged that the cooling load is likely to change with the seasons depending on the geographical location. However, for the case of this study, both computing and cooling loads are assumed to be constant.

Energy Storage

The battery was sized to meet the energy needs of the techniques proposed in this paper. A large battery bank was necessary because a lead acid battery bank should only be discharged to 50% of its maximum capacity [25]. The battery bank capacity should be close to twice the average solar energy generation in a given day during a summer month, when solar energy generation is at its highest. Also, lead acid battery specifications are typically given assuming that the batteries are discharged over a 20 hour period. If the end user discharges the lead-acid battery faster than a 20 hour rate, the actual capacity of the battery is less than the specified capacity [26]. The irregular discharging conditions in this paper called for a high capacity battery bank, which could be achieved by wiring several 2V batteries with high energy capacity in series and parallel. The Crown 2CRP3690, rated at 2 Volts and 2550 Amp-Hours at a 20 hour rate, was used in the model [27]. The batteries could be wired in series to reach up to 60VDC, thus eliminating the need to add excessive parallel strings to the battery bank, which damages the batteries [28]. This can be shown by dimensional analysis. The battery bank energy storage can be calculated using Equation 2:

$$E = V_{Batt} \times Capacity \quad \textbf{(Equation 2)}$$

Where the capacity is in units Amp*hr, the battery voltage is in units volts, and the energy in Watt*hr. The usable power of the battery system is different when discharged at different current rates, so this was accounted for by applying Peukert's Law, which computes the usable capacity of lead acid batteries when discharged at certain rates [26].

Hybrid Management Based on Pricing Models

Two pricing models, Time of Use agreement, and Coincident Peak agreement were considered and used to develop energy management strategies using hybrid energy storage of renewable energy.

Time of Use Agreement

A Time of Use agreement is a specific way that a utility company charges its customers for electricity. In a Time of Use agreement, there is something known as On-Peak hours. During On-Peak hours, customers are charged more for electricity versus Off-Peak hours. These hours are set forth by individual utility companies, and the specific times are dependent upon the demand that the utility company receives. Time of Use agreements do not take into account real time pricing, meaning that the On-Peak hours are static and openly discussed when the contract is signed. All electricity agreements involve a maximum demand charge. However, in Time of Use agreements, utility companies may choose to only charge a maximum demand charge during the On-Peak hours. Some companies may still charge their customers for both the On-Peak maximum demand and the overall maximum demand for a given month.

In the research presented in this paper, the Time of Use agreement was done with Florida Power & Light (FPL) data as seen in Table 1. The hybrid energy management allows the end user to store the solar energy until On-Peak hours. Then, the demand shaving approach is applied during On-peak hours to lower the maximum On-Peak demand charge [29]. By shaving the peaks, or lowering the maximum On-Peak demand, end users could save $8.28 per kW of maximum On-Peak demand shaved. The variability of the system design depends on the battery bank capacity, which must have the ability to safely discharge a day's worth of solar energy within an 8 to 9 hour frame for this particular scenario. Since the On-Peak hours change semiannually (5 months in the winter and 7 months in the summer) the technique was applied to two different months, January and July.

Florida Power & Light Time of Use Agreement (GSLDT-1) Parameters [29]			
Time of Year	**On-Peak Hours**	**On Peak Demand Rate**	**Off Peak Demand Rate**
November 1 - March 31	6 A.M. to 10A.M, 6 P.M. to 10 P.M.	$8.28 per kW	none
April 1 - October 31	12P.M. to 9P.M.		

Table 1: FPL Time of Use Agreement Schedule Parameters

The system needed 72 batteries because of the technique employed. During the month of July, the average solar energy generated was 148.388kWh per day. The battery bank had to have a capacity of at least 296.776kWh because a battery bank should only be discharged to 50% of its maximum capacity [25]. Additionally, the On-Peak demand shaving required the batteries to discharge the solar energy over a period of approximately six hours. Peukert's Equation was applied to account for a decrease in capacity because of the short discharge period [26, 30].

Coincident Peak Agreement

Unlike Time of Use agreements, Coincident Peak agreements do not have On-Peak hours, but rather, one Coincident Peak. A Coincident Peak is one hour out of every month when the utility company has the most demand for electricity. With Coincident Peak pricing agreements, the Coincident Peak is not known until the end of the month. However, utility companies that charge a Coincident Peak typically provide predictions for Coincident Peaks up to 24 hours in advance [31]. During a Coincident Peak, customers are charged an extremely high electricity rate for energy consumption during that hour. These companies also charge a maximum facility demand charge.

For Coincident Peak pricing agreements, this paper used the 'E-400' schedule from Fort Collins Utilities (FCU), as seen in Table 2. According to Fort Collins Utilities, Coincident Peak charges alone account for approximately 23% of their customers' electricity bills [32].
Therefore, a photovoltaic energy management approach, which reduces the Coincident Peak demand, would have optimal operational cost reduction. Since the coincident peak charge varies during different seasons, the technique was applied to two months, January and July.

Fort Collins Utilities Coincident Peak Agreement (E-400) Parameters [33]			
Time of Year	**On-Peak Hours**	**Coincident Peak Rate**	**Off Peak Rate**
June -July	1 hr/month	$11.01 per kWh	$0.040/kWh
Aug-May	1 hr/month	$8.15 per kWh	$0.039/kWh

Table 2: FCU Coincident Peak Agreement Schedule Parameters

The Coincident Peak agreement required a system that utilized 120 batteries. In July, the average solar energy generation was 180.921 kWh per day, thus battery bank had to have a capacity size of approximately 360 kWh. The proposed technique required the batteries to discharge solar energy in one hour. According to Peukert's equation, the battery bank had to be oversized to meet the rapid discharge requirements [30].

3: RESULTS & DISCUSSION

Both hybrid management strategies for the two pricing agreements are considered and specific set points are solved for, which are used in managing the use of the renewable energy stored in the battery. The result is then evaluated from a return on investment perspective.

3.1: Brown Energy Reduction

Time of Use Agreement

For the Time of Use agreement, peak shaving techniques provided the optimal cost reduction in the data center. An optimization algorithm was developed in the MATLAB computing environment, and applied to two seasonal pricing schemes.

In Figure 4, the graph indicates that the On-Peak hours peak for the month of January was effectively shaved, as indicated by the On-Peak power cap. By capping the maximum grid energy drawn by the data center during the On-Peak hours, the data center can keep the maximum demand charge as low as possible. Whenever the data center needs to draw more power than the On-Peak power cap, the accumulated solar energy stored within the batteries is discharged to provide the extra power for the data center load. For Figure 4, the On-Peak power cap was approximately 1,104.365 kW. To calculate this value, the study applied Equation 3 to an algorithm in order to approximate the On-Peak power cap:

$$\left[\int_{t=A}^{t=10} P(t) - P_{cap} \quad + \quad \int_{t=18}^{t=B} P(t) - P_{cap}\right] \quad - \quad E_{solar}$$

(Equation 3)

where, P(t) is the data center load curve, P_{cap} is the y value of the On-Peak power cap, A is the point where P_{cap} intersects P(t) on the interval t=6 to t=10, B is the point where P_{cap} intersects P(t) on the interval t=18 to t=22, and E_{solar} is the average total photovoltaic energy generation in a day of that month.

For the algorithm, the study set the initial value of the constant P_{cap} as the maximum y-value of the Coincident Peak hours. In this scenario, the initial P_{cap} is P(18). As the algorithm runs, each iteration decreases the value of P_{cap} by 0.001kW until the equation above equals zero, with a tolerance < 0.01kW. By iterating through thousands of samples, the algorithm offers an accurate approximation of the optimal On-Peak power cap.

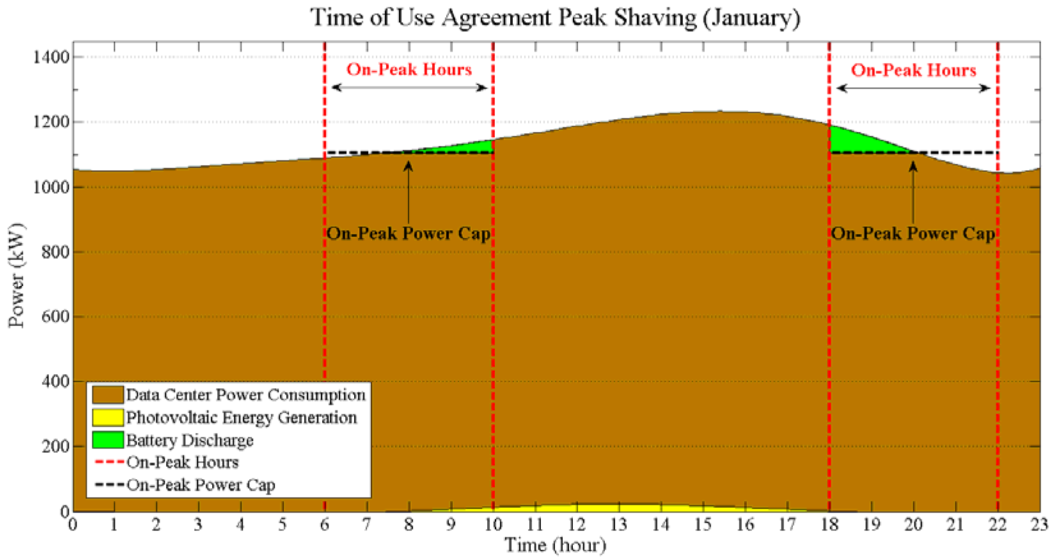


Figure 4 : Florida Power & Light Time of Use agreement On-Peak demand load shaving for January

In Figure 5, the graph represents the same Time of Use pricing agreement, but under different On-Peak hours. Unlike the month of January, the On-Peak hours for July are contained within one long time interval. The On-Peak Power Cap for July was 1190.612 kW. To get this value, the previous equation was adapted to fit the new On-Peak hours (Eq. 4):

$$\left[\int_{t=A}^{t=B} P(t) - P_{cap}\right] \quad - \quad E_{solar}$$ **(Equation 4)**

where, P(t) is the data center load curve, P_{cap} is the y value of the On-Peak power cap, A is the point where P_{cap} intersects P(t) on the interval t=12 to t=15.393(absolute max), B is the point where P_{cap} intersects P(t) on the interval t=15.393(absolute max) to t= 21, and E_{solar} is the average total Photovoltaic energy generation in a day of that month.

For the algorithm, the same concepts as the algorithm for the month of January was applied. However the initial value of P_{cap} was set as the absolute maximum of P(t) since the maximum y-value of the On-Peak hours happens to be the absolute maximum of the entire load curve. After solving for the critical points, the absolute maximum of the function is determined to be at approximately t=15.393. Then, I ran the algorithm to approximate the On-Peak power cap of July by solving for the zero of the equation, with a tolerance < 0.01.

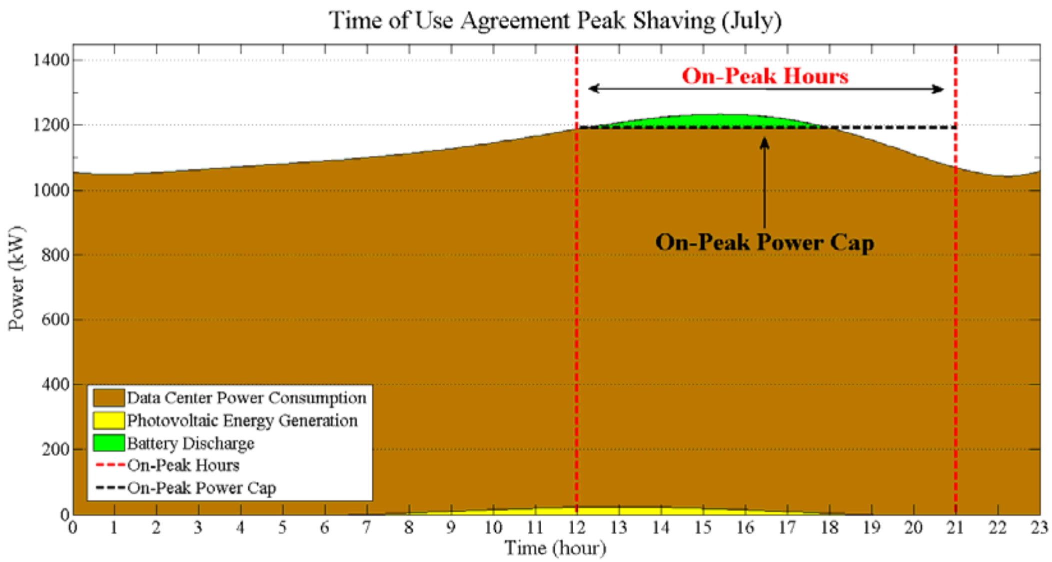


Figure 5: Florida Power & Lights Time of Use agreement On-Peak demand load shaving for July

Coincident Peak Agreement

The coincident peak load shaving technique is applied to two months, January and July.

In Figure 6, the photovoltaic energy is discharged within one hour - the Coincident Peak. Here, this study caps the brown energy consumption during the Coincident Peak hour. If the data center draws more power than the Coincident Peak power cap, the deficiency is supplied by the battery powered by the solar energy. For the month of January, the Coincident Peak power cap was set at 1,019.157 kW. To get this value, the study applied Equation 5a:

$$\left[\int_{t=18}^{t=19} P(t) - P_{cap}\right] \quad - \quad E_{solar}$$

(Equation 5a)

where, $P(t)$ is the data center load curve, P_{cap} is the y value of the Coincident Peak power cap, and E_{solar} is the average total Photovoltaic energy generation in a day of that month.

The algorithm approximates the zero of the equation. P_{cap} is initially set at the maximum y-value of the Coincident Peak hour. For January, this initial value was set at P(18). The algorithm continuously samples thousands of P_{cap} values until the desired root of the equation was within 0.01 kW.

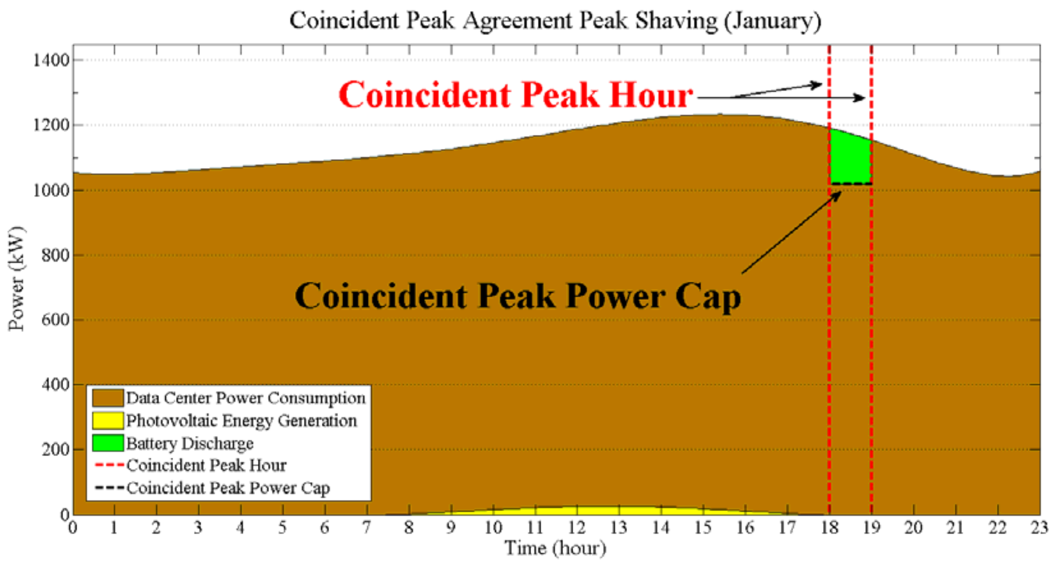


Figure 6: Fort Collins Utility Coincident Peak agreement load shaving for January

Figure 7 shows the same technique applied to the month of July. In July, the Coincident Peak was set between 4 P.M. and 5 P.M. Although similar to January, the key differences lie between the Coincident Peak hour and the amount of solar energy generated. For July, the Coincident Peak power cap was set at 1,044.913 kW. The same equation from January was applied, but the upper and lower bounds were adjusted according to the Coincident Peak hours (Eq. 5b).

$$\left[\int_{t=16}^{t=17} P(t) - P_{cap}\right] \quad - \quad E_{solar}$$ **(Equation 5b)**

where, P(t) is the data center load curve, P_{cap} is the y value of the Coincident Peak power cap, and E_{solar} is the average total Photovoltaic energy generation in a day of that month. The algorithm was reused from January, where the initial value of P_{cap} is set as P(16).

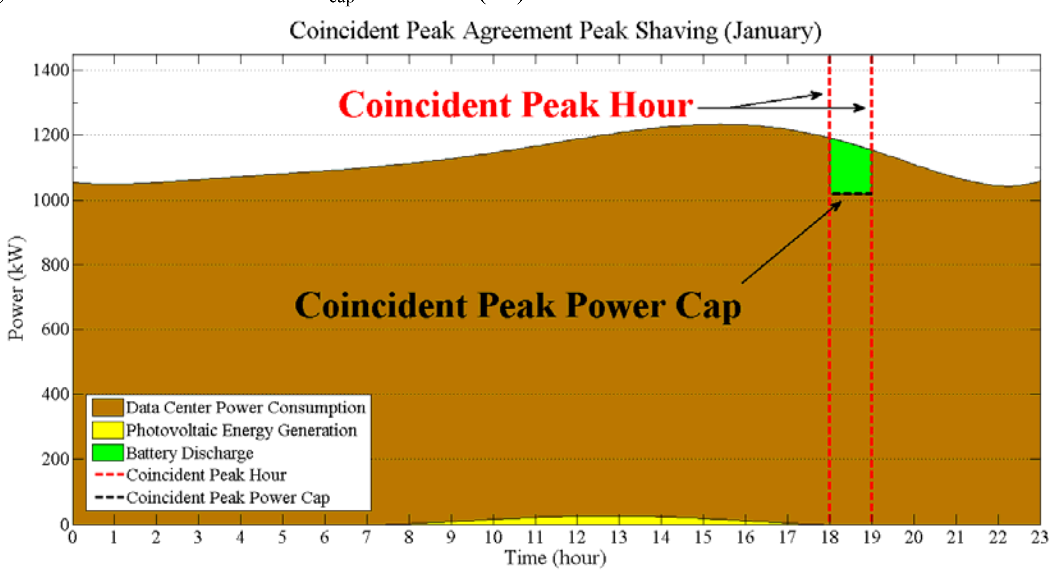


Figure 7: Fort Collins Utility Coincident Peak agreement load shaving for January

3.2: Return on Investment

The return on investment (ROI) for hybrid Photovoltaic systems were variable. End users have to take into account pricing agreements, and closely examine the price rates of their respective utility provider. Another factor that played a key role in the ROI was the system configurations. Certain Photovoltaic energy management techniques called for more expensive systems to carry out the proposed technique.

Time of Use

The Time of Use Agreement used in the paper was Florida Power & Light rate schedule 'GSLDT-1.' The grid power consumption was capped at a certain value during the On-Peak hours. Depending on where the On-Peak power cap was set, this study was able to reduce energy costs in both January and July. To get a grasp of the total cost of the system, this paper focused on the grand total of the system cost for the Time of Use agreement (Table 3). The total cost of the system amounts to $151,648.

Time of Use Operational Cost Reduction					
Cost Reduction [29]		**Capital Cost**		**Return on Investment (ROI)**	
Maximum On-Peak Demand (January)	$723.59	Solar Modules [22]	$59,000.00	$/Year	$7,211.71
Base Energy (January)	$86.12	Batteries [27]	$98,640.00		
Maximum On-Peak Demand (July)	$356.01	Installation & Associated Parts [34]	$59,000.00	%/Year	4.76%
Base Energy (July)	$95.87	Federal Incentive [35]	($64,992.00)		
Annual	$7,211.71	Total	$151,648.00	Break Even Point (BEP)	21.03 Years

Table 3 : Florida Power & Light's Time of Use agreement cost analysis

Time of Use agreements did not show much potential for a practical return on investment. During January, operational costs were $809.71 less. During July, there was a cost reduction of $451.88. It was assumed that these cost reductions were indicative of the 5 "winter" months and 7 "summer" months, and an annual cost savings of $7,211.71 was made. To make a positive return on investment, it would take approximately 21 years.

Coincident Peak

The Coincident Peak agreement used in the paper was Fort Collins Utilities rate series 'E-400' [33]. The study capped the grid power consumption at a certain value during the Coincident Peak hour. Depending on where the Coincident Peak power cap was set, the data center was able to reduce costs in both January and July (Table 4).

Coincident Peak Operational Cost Reduction					
Cost Reduction [33]		**Capital Cost**		**Return on Investment (ROI)**	
Coincident Peak (January)	$1,406.67	Solar Modules [22]	$59,000.00	$/Year	$21,200.70
Base Energy (January)	$188.55	Batteries [27]	$164,400.00		
Coincident Peak (July)	$2,054.15	Installation & Associated Parts [34]	$59,000.00	%/Year	10.72%
Base Energy (July)	$227.09	Federal Incentive [35]	($84,720.00)		
Annual	$21,200.70	Total	$197,680.00	Break Even Point (BEP)	9.32 Years

Table 4: Fort Collins Utility Coincident Peak Agreement cost analysis

Despite the major increase in system cost in comparison to that of the Time of Use agreement, the Coincident Peak agreement allowed a practical return on investment. During the month of January, $1,595.22 was saved. For July, $2,281.24 was saved. Assuming that the cost reduction for January represented every non summer month and the cost reduction for July represented all three summer months, there is an annual grid energy cost reduction of $21,200.70, making a return on investment in just over 9 years.

3.3: DISCUSSION

This study tried to show the applicability and practicality of hybrid photovoltaic systems in data centers. Results suggest that reaching a practical return on investment is heavily dependent on uncontrollable factors such as pricing scheme, irradiance at certain geographical locations, and the relative cost of electricity for a given utilities provider.

Among the two pricing agreements, the Coincident Peak agreement showed much greater energy cost reductions, up three times as much, for the techniques used in this paper. This suggests that Time of Use agreements may be at a disadvantage when investing in hybrid photovoltaic systems. There were variations in energy cost reductions throughout different times of the year. Results indicated that on average, one season would see nearly twice the energy cost reductions of another. For the Time of Use agreement, the most energy cost reductions were made in the winter months while Coincident Peak agreement experienced the opposite.

The hybrid photovoltaic system cost varied between the two pricing agreements. The Time of Use agreement system had a price that was expected of a hybrid system of that size. The price of the Coincident Peak agreement system was nearly 125% of the cost of the Time of Use system. Coincident peak power capping required a much larger battery bank capacity than On-Peak demand shaving. Despite the difference in system costs, the more expensive Coincident Peak system saw a nine year return on investment whereas the cheaper, Time of Use system, saw a twenty one year return on investment. The Coincident Peak system involved a higher capital investment, but the ratio of energy cost reduction per year to the initial system cost was much greater for the Coincident Peak agreement than that of the Time of Use system.

This study's results suggest that the practicality of hybrid photovoltaic integration in data centers may be more complex than just cost versus cost reduction. Prior to this study, the author did not come across any papers that proposed, conducted, and evaluated techniques of hybrid photovoltaic integration in data centers that made an in-depth analysis of practicality and financial returns among different pricing agreements. The author believes that the study conducted in this paper marks the initial stages of research that emphasizes practical integration of photovoltaics in data centers.

4: CONCLUSIONS AND FUTURE WORK

4.1: Closing Remarks

Coincident Peak pricing agreements offered the most operational cost reductions and fastest return on investment, despite a high system cost. Time of Use pricing agreements offered less operational cost reductions and slower return on investment, despite a lower system cost. The variability of cost reductions throughout seasons and the unique pricing schedules among utilities providers played a key role in the overall cost analysis. Results suggest that hybrid photovoltaic integration in data centers may be a practical solution to operational cost and brown energy consumption reduction in geographical locations where the techniques proposed in this paper integrate well with the utilities provider's rate schedule.

4.2: Future Work

Although the techniques proposed in this paper seemed to be effective in some cases, further research should be done in optimizing and developing an approach that fits all pricing schemes. A more efficient and universal algorithm may be devised by applying concepts of the bisection method, and an in-depth cost analysis of storage devices should be made to achieve a more practical return on investment.

5: REFERENCES

[1] U.S. Department of Energy, "Data Center Energy Consumption Trends," [Online]. Available: http://www1.eere.energy.gov/femp/program/dc_energy_consumption.html. [Accessed 1 July 2013].

[2] U.S. Environmental Protection Agency ENERGY STAR Program, "Report to Congress on Server and Data Center Energy Efficiency," EPA, 2007.

[3] U.S. Energy Information Administration, "Frequently Asked Questions," EIA, [Online]. Available: http://www.eia.gov/tools/faqs/faq.cfm?id=97&t=3. [Accessed 1 July 2013].

[4] J. G. Koomey, "Growth in Data Center Electricity Use 2005 to 2010," Analytics Press, Burlingame, CA, 2011.

[5] Google, "Efficiency," [Online]. Available: http://www.google.com/about/datacenters/efficiency/. [Accessed 2 July 2013].

[6] Google, "Renewable Energy," [Online]. Available: http://www.google.com/about/datacenters/renewable/. [Accessed 2 July 2013].

[7] Apple, "Apple and the Environment," [Online]. Available: http://www.apple.com/environment/renewable-energy/. [Accessed 2 July 2013].

[8] Open Compute Project, "Energy Efficiency," [Online]. Available: http://www.opencompute.org/about/energy-efficiency/. [Accessed 2 July 2013].

[9] M. Stansberry and J. Kudritzki, "Uptime Institute 2012 Data Center Industry Survey," Uptime Institute, New York, 2012.

[10] National Renewable Energy Laboratory, "Get Your Power from the Sun," U.S. Department of Energy, Washington, D.C., 2003.

[11] S. Akoush *et al.*, "Free Lunch: Exploiting Renewable Energy for Computing," in *HotOS*, 2011.

[12] C. Li *et al.*, "SolarCore: Solar Energy Driven Multi-core Architecture Power Management," in *IEEE High Performance Computer Architecture (HPCA)*, 2011.

[13] K. Le *et al.*, "Cost-and Energy-Aware Load Distribution Across Data Centers," in *HotPower*, 2009.

[14] Í. Goiri *et al.*, "GreenSlot: Scheduling Energy Consumption in Green Datacenters," in *ACM International Conference for High Performance Computing, Networking, Storage and Analysis*, 2011.

[15] G. Ghatikar *et al.*, "Demand Response Opportunities and Enabling Technologies for Data Centers: Findings from Field Studies," Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA, 2012.

[16] M. Ghamkhari and M.-R. Hamed, "Optimal Integration of Renewable Energy Resources in Data Centers with Behind-the-meter Renewable Generator," in *IEEE International Conference on Communications (ICC)*, 2012.

[17] Emerson Power Network, "Energy Logic: Reducing Data Center Energy Consumption by Creating Savings that Cascade Across Systems," Emerson Electric Co., Columbus, OH, 2009.

[18] Basler Electric Company, "Introduction to Synchronizing," Basler Electric Company, Highland.

[19] Nation Renewal Energy Laboratory, "PVWatts," NREL.

[20] PV Performance Modeling Collaborative, "Inverter Efficiency," [Online]. Available: http://pvpmc.org/modeling-steps/dc-to-ac-conversion-2/inverter-efficiency/. [Accessed 10 July 2013].

[21] J. W. Stevens and G. P. Corey, "A Study of Lead-acid Battery Efficiency Near Top-of-charge and the Impact on PV System Design," in *IEEE Conference Record of the Twenty Fifth Photovoltaic Specialists Conference*, 1996.

[22] Wholesale Solar, "Suniva MVX 250-60-5-701," Wholesale Solar, [Online]. Available: http://www.wholesalesolar.com/products.folder/module-folder/Suniva/MVX-250-60-5-701.html. [Accessed 8 July 2013].

[23] G. Albright *et al.*, "A Comparison of Lead Acid Lithium-ion in Stationary Storage Applications," AltEnergy eMagazine, 2012.

[24] National Renewable Energy Laboratory, "Battery Power for Your Residential Solar Electric System," U.S. Department of Energy, 2002.

[25] SmartGauge Electronics, "The 50% Rule for Deep Cycle Batteries," [Online]. Available: http://www.smartgauge.co.uk/50percent.html. [Accessed 20 July 2013].

[26] BatteryStuff, "Determine Run Time for Specific Load," 22 February 2012. [Online]. Available: http://www.batterystuff.com/kb/tools/calculator-for-load-specific-run-time.html. [Accessed 21 July 2013].

[27] Northern Arizona Wind & Sun, "Crown 2CRP3690 Renewable Power Series," Crown, [Online]. Available: http://www.solar-electric.com/repoba2vo25a.html. [Accessed 24 7 2013].

[28] C. Freitas, "High-Capacity Battery Banks:," SolarPro, 2012.

[29] Florida Power & Light Company, "Rate Schedules," FPL.

[30] C. Gibson, "Peukert Equation Calculator," SmartGauge Electronics.

[31] Z. Liu *et al.*, "Data Center Demand Response: Avoiding the Coincident Peak via Workload Shifting and Local Generation," in *ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, 2013.

[32] Fort Collins Utilities, "Coincident Peak," [Online]. Available: http://www.fcgov.com/utilities/business/rates/electric/coincident-peak. [Accessed 11 July 2013].

[33] Fort Collins Utilities, "Large Commercial/Industrial," [Online]. Available: http://www.fcgov.com/utilities/business/rates/electric/e400. [Accessed 11 July 2013].

[34] Advanced Research Projects Agency-Energy, "$1/W Photovoltaic Systems," U.S. Derpartment of Energy.

[35] Database of State Incentives for Renewables & Efficiency, "Business Energy Investment Tax Credit (ITC)," [Online]. Available: http://www.dsireusa.org/incentives/incentive.cfm?Incentive_Code=US02F&re=1&ee=1. [Accessed 15 July 2013].

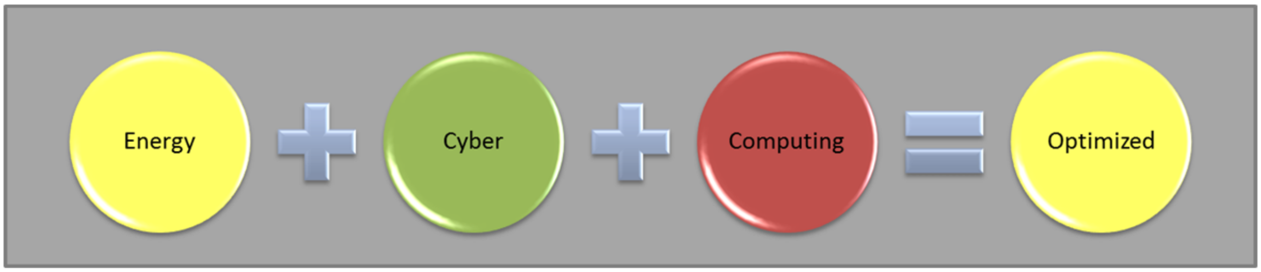
APPENDIX D: Dashboard Widgets

FY13 DEMO WIDGETS

For use in demonstration of what is possible with the EFFIS-INT dashboard. Still conceptual but illustrates how it can be integrated with Ozone Widget Framework.

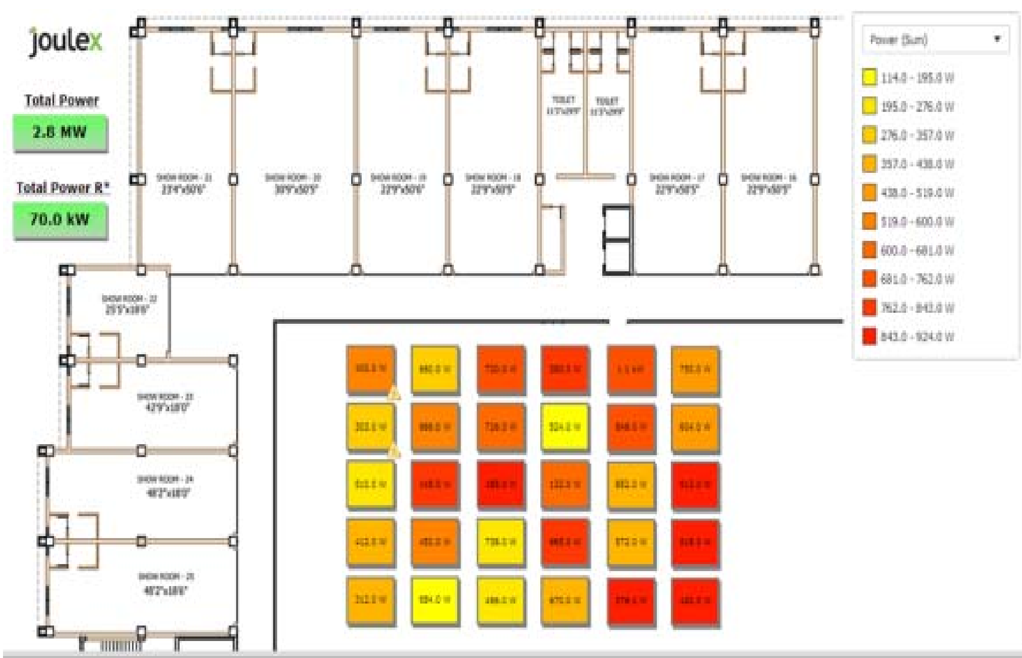
The following are rough concepts only with regards to the individual widgets. The widgets in the final design will need to interact with each other and respond to changes or examination of changes applied to one widget will update or modify related or corresponding widgets.

STATUS SUMMARY WIDGET



- Quick look color coded summary of status of main energy components.
- Quick look color coded summary of the current optimization of energy.
- Optimization levels for all main components add up to the total optimized level

SYSTEM MAP WIDGET



- This widget shows a physical representation of the energy related
- Hardware along with color coded representation of their state.

ALERTS WIDGET

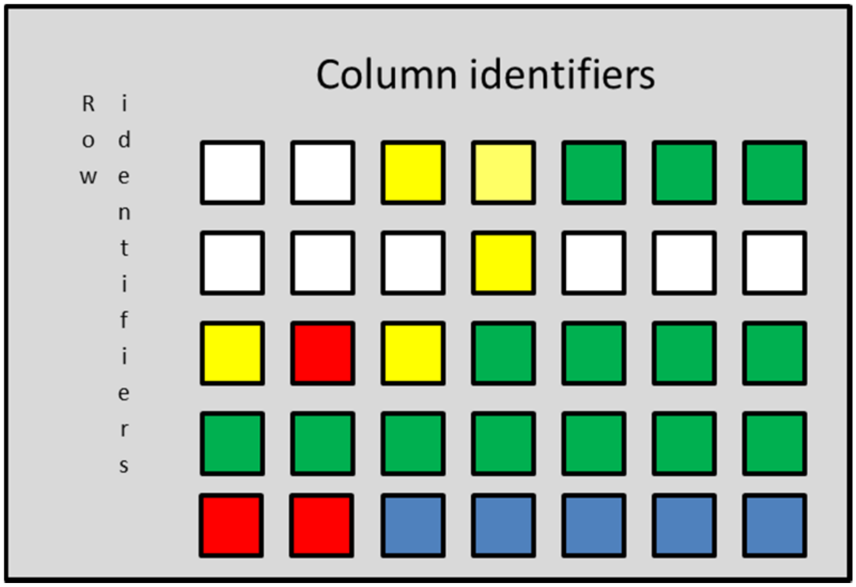
- Text summary of what alert is about including action if needed.
- Background coloring indicates the severity of the alert.
- Color coding needs to be identified including definition of alerts.
- Could contain an associate icon which represent the area of the alert.
- User can set parameters of the scrolling alerts such as speed, colors, and other.
- Enhancement could include mouse rollover to get a tooltip of amplifying information.
- Enhancement of clicking on item brings up a window with information about the alert.
- Can include color, icon or other coding to indicate a user's input is required.
-

TASK TRACKER WIDGET

Task Tracker

Title	Due Date	Assigned To
Check power supply	11/2/13	Christian
Install new fans	4/21/13	Bryan

- Table based task tracker
- As a simple means to see
- Tasks in a tabular format.

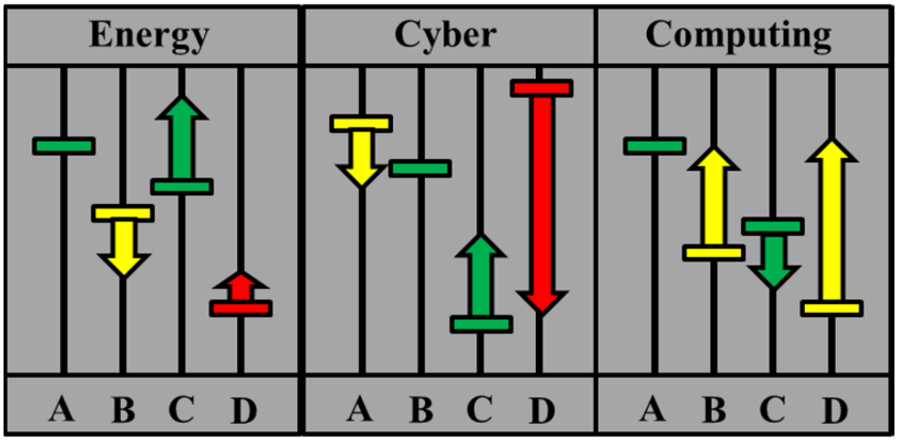


- Simple icon/symbol based tasks and status of task. Each symbol/icon represents a task and the task state. They can be categorized by rows and column headers.

BASIS OF ASSESSMENT WIDGET

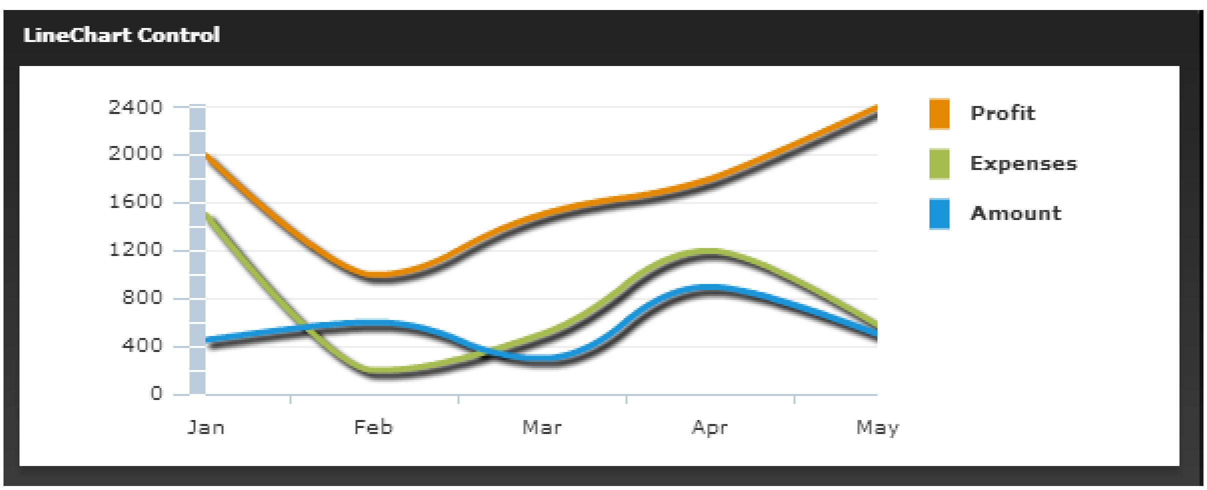
- This is an example of a Basis of Assessment widget used in the MMWS project.
- Left hand column is color coded list of status of specific systems.
- Right hand column lists specific data items that provide overall color code in column
- Middle section provides a time based status of right hand column data items.
- Middle section if time is applicable shows how status changes over time.
- Right hand column will color code to current status in time (current time).
- Time line goes from current time and back in history.
- If widget is linked to future predictions then then the user should be able to scroll forward
- In time.

COURSE OF ACTION WIDGET



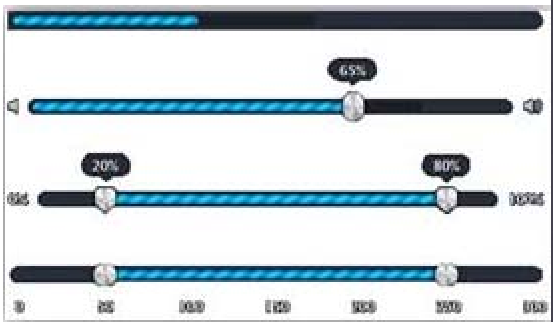
- Conceptual interface showing columns of potential changes in 3 main categories.
- Each column is a given specific item that can be changed in that category.
- Bar represents the current level. Bar color indicates current level acceptance.
- Arrow represents the course to taken with respect to current bar level.
- Color coding is conceptual and needs further definition but may indicate the
- Confidence of the course of action to be taken.

MONITOR POWER WIDGET



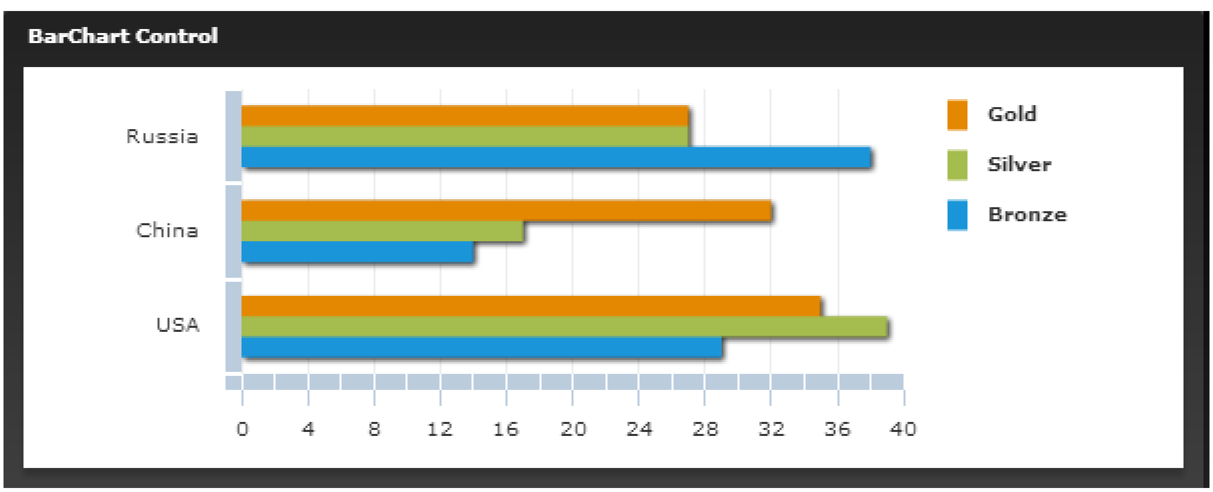
- A basic line chart can be utilized to show power usage for various components
- Over time. Here the values would be changed to correspond to power usage
- Levels and power usage components in the system. Other basic charts types
- Could also be utilized. A time slider can also be inserted to view specific
- Details on a given period of time or view larger view.

POWER CONTROL WIDGET



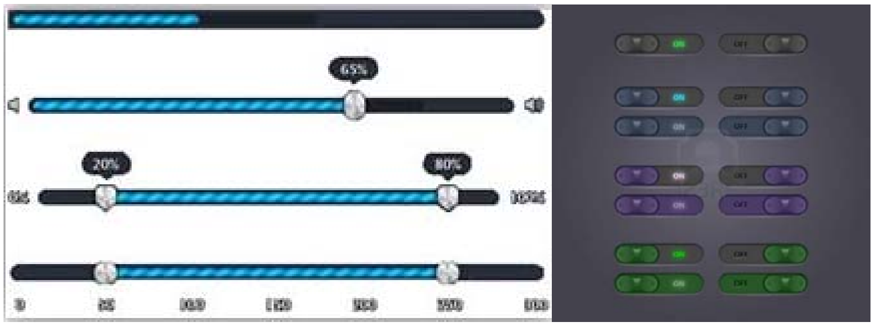
- Control widgets in general allow the user to modify current settings.
- Changes can be user based decision or based on course of action recommendations.
- Change can be applied via slider controls for just the setting or setting high and low limits.
- Change can be applied if applicable via on or off switch type interface.
- Due to the nature of control changes a user authorization must occur.
- Each change must be verified by the user in order for it to take place.

MONITOR COOLING WIDGET



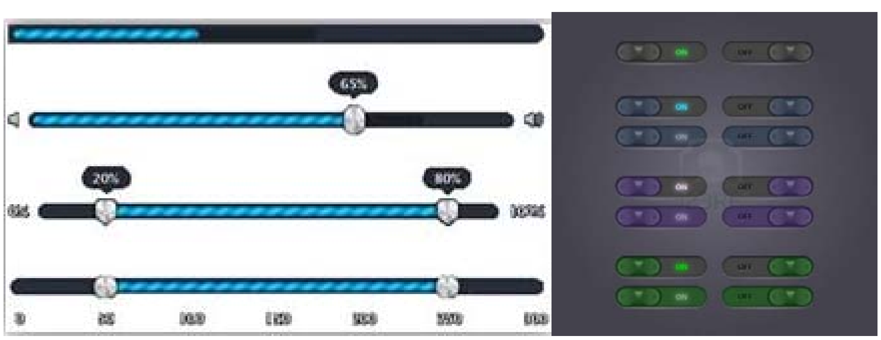
- A basic bar chart can be utilized to show cooling for various components
- Over time. Here the values would be changed to correspond to computing
- Levels applied to virtual machines in the system. Other basic charts types
- Could also be utilized. A time slider can also be inserted to view specific
- Details on a given period of time or view larger view. Line or other chart
- Type can be used.

COOLING CONTROL WIDGET



- Control widgets in general allow the user to modify current settings.
- Changes can be user based decision or based on course of action recommendations.
- Change can be applied via slider controls for just the setting or setting high and low limits.
- Change can be applied if applicable via on or off switch type interface.
- Due to the nature of control changes a user authorization must occur.
- Each change must be verified by the user in order for it to take place.
- MONTIOR COMPUTING WIDGET
- A basic area chart can be utilized to show power usage of computering systems.
- Over time. Here the values would be changed to correspond to computing
- Levels applied to components in the system. Other basic charts types
- Could also be utilized. A time slider can also be inserted to view specific
- Details on a given period of time or view larger view. Line or other chart
- Type can be used.

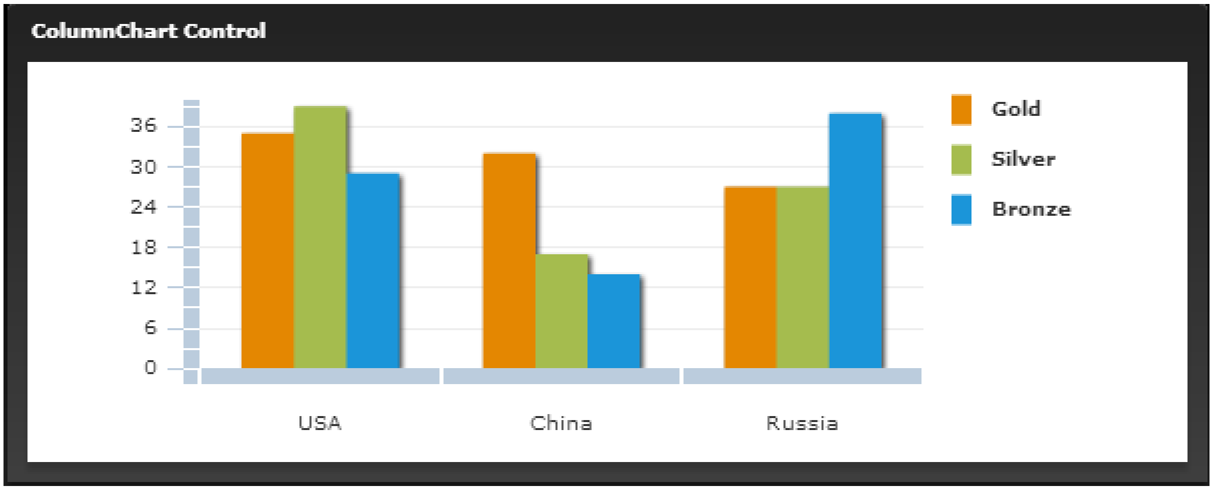
COMPUTING CONTROL WIDGET



- Control widgets in general allow the user to modify current settings.
- Changes can be user based decision or based on course of action recommendations.
- Change can be applied via slider controls for just the setting or setting high and low limits.
- Change can be applied if applicable via on or off switch type interface.

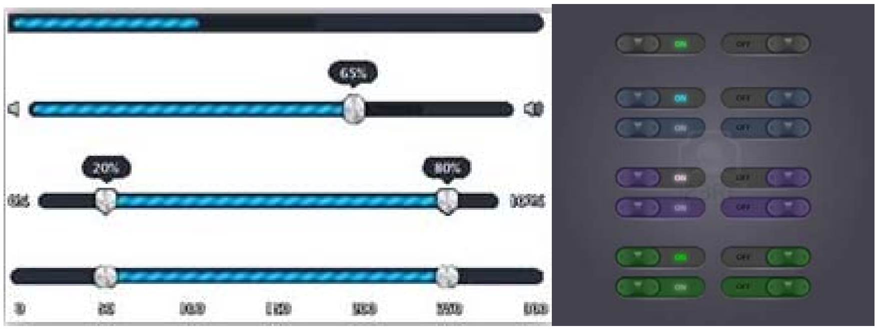
- Due to the nature of control changes a user authorization must occur.
- Each change must be verified by the user in order for it to take place.
-

MONITOR CYBER SECURITY WIDGET

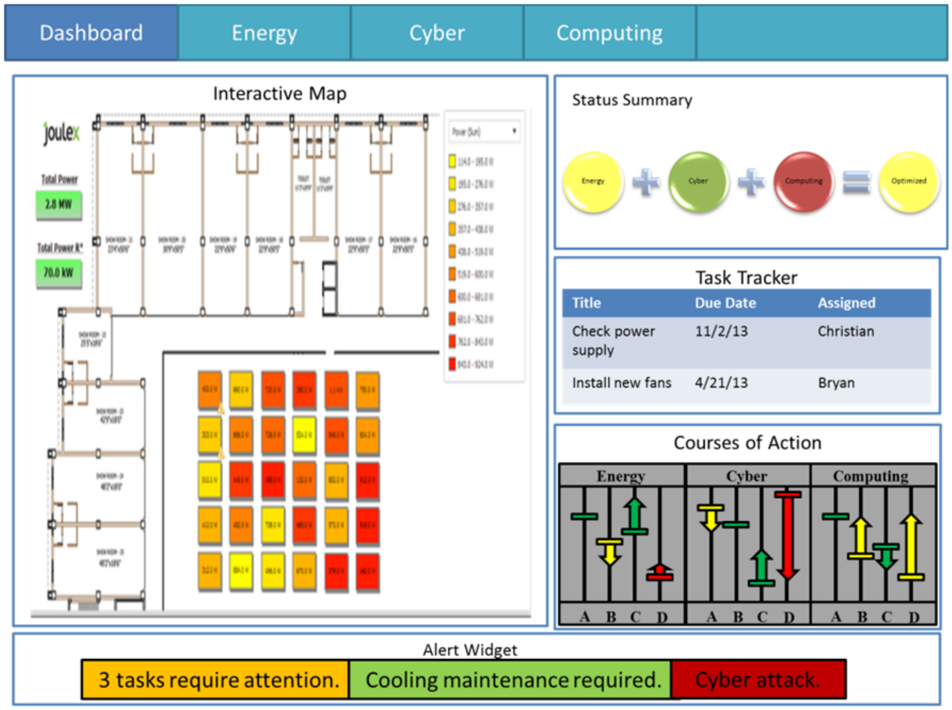


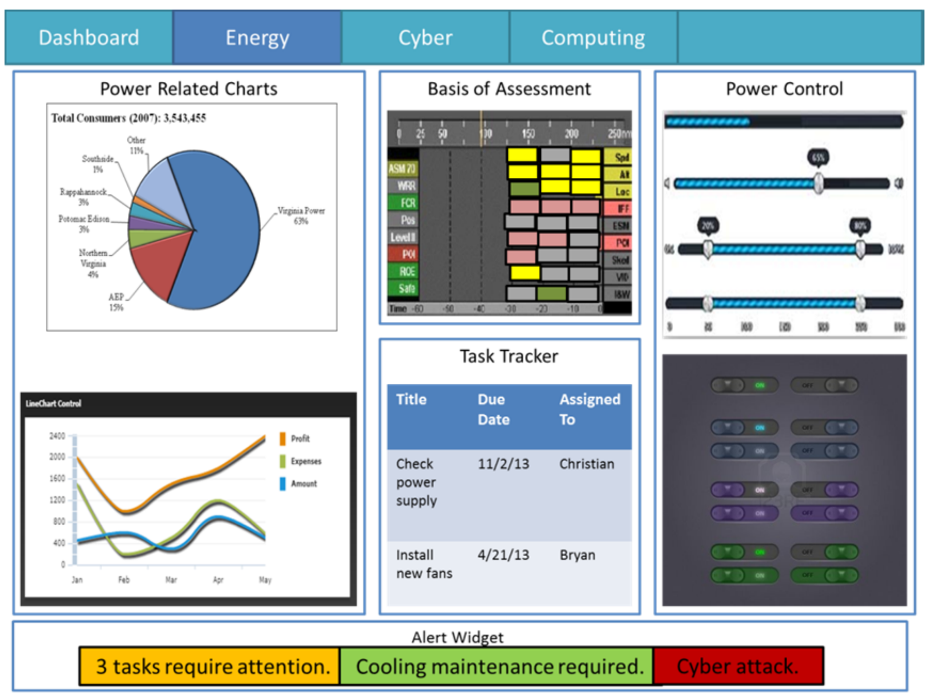
- A basic bar chart can be utilized to show cyber security for various components.
- Here the values would be changed to correspond to cyber security
- Levels applied to components in the system. Other basic charts types
- Could also be utilized. A time slider can also be inserted to view specific
- Details on a given period of time or view larger view. Line or other chart
- Type can be used.

CYBER SECURITY CONTROL WIDGET



- Control widgets in general allow the user to modify current settings.
- Changes can be user based decision or based on course of action recommendations.
- Change can be applied via slider controls for just the setting or setting high and low limits.
- Change can be applied if applicable via on or off switch type interface.
- Due to the nature of control changes a user authorization must occur.
- Each change must be verified by the user in order for it to take place.


Dashboard
Energy
Cyber
Computing
Interactive Map
joulex
Total Power
2.8 MW
70.0 kW
Status Summary
Energy
Cyber
Computing
Optimized
Task Tracker
Title
Due Date
Assigned
Check power supply
11/2/13
Christian
Install new fans
4/21/13
Bryan
Courses of Action
Energy
Cyber
Computing
A B C D
A B C D
A B C D
Alert Widget
3 tasks require attention.
Cooling maintenance required.
Cyber attack.


Dashboard
Energy
Cyber
Computing
Power Related Charts
Total Consumers (2007): 3,543,455
Other 11%
Southside 1%
Rappahannock 3%
Potomac Edison 3%
Northern Virginia 4%
AEP 15%
Virginia Power 63%
LineChart Control
Profit
Expenses
Amount
Jan
Feb
Mar
Apr
May
Basis of Assessment
Task Tracker
Title
Due Date
Assigned To
Check power supply
11/2/13
Christian
Install new fans
4/21/13
Bryan
Power Control
Alert Widget
3 tasks require attention.
Cooling maintenance required.
Cyber attack.

APPENDIX E: Software Requirements

Software Requirements Specification

Version 1.0

August 14, 2013

EFFIS-INT Dashboard

by Bryan Croft

SSC Pacific
Science and Technology
Capability Investment Project
Energy Optimization
For Command and Control Systems

Table of Contents

List of Figures

1.0. Introduction

1.1. Purpose

This document contains a detailed description of the EFFIS-INT Dashboard requirements as a user interface for Command and Control (C2) systems. It will explain the purpose and features of the system, the interfaces of the system, what the system will do, the constraints under which it must operate and how the system will react to external stimuli. This document is intended for both the stakeholders and the developers of the system and will be used to design and develop the EFFIS-INT dashboard.

1.2. Scope of Project

The EFFIS-INT dashboard software system is intended to be a visualization interface of optimized energy usage for C2 systems. This system will be designed to display relative energy optimization content that is provided by examining the energy related components of C2 systems. In addition it will provided unique features to allow the end user to monitor the energy usage while providing courses of action alternatives from the current actions. Included in this are displays showing the basis of the assessments made by analysis systems the look for patterns and indications leading to energy based course of action.

More specifically, this system is designed to allow the end user to manage and control the energy usage of the system with support from energy data analysis tools. The software will facilitate the end users in making decisions which help to optimize the usage of energy as well manage the C2 computing system to maintain operational requirements of those systems. A novel approach to the EFFIS-INT effort is also to determine if energy related data can provide indication of cyber threats.

1.3. Glossary

Term	Definition
C2	Command and Control
EFFIS-INT	Energy Focus Fused Information System - Integration
Dashboard	Display interface containing various sub components which illustrated summarized information sets.
End User	A person who is a primary user of an application or interface to a system and is generally required to interact with the system.
OWF	Ozone Widget Framework a web based framework which contains, layouts, and provided interaction between widgets or small applications.
Software Requirements Specification	A document that completely describes all of the functions of a proposed system and the constraints under which it must operate. For example, this document.
Stakeholder	Any person with an interest in the project who is not a developer.

1.4. References

1) Croft, B., "Software Architecture Document (SAD) EFFIS-INT Integrated C2 Dashboard Architecture ", December 2012, SSC Pacific EFFIS-INT Project Documentation.
2) Hoff, T., Hauser, A., "Applying A Cognitive Engineering Approach to Interface Design of Energy Management Systems", PsycholNology Journal, 2008, Volume 6, Number 3, pages 321 - 345.
3) "A Guide to Creating Dashboards People Love to Use", Novemeber 2009, Publicationof Juice, Inc.
4) Haves, Philip, "Building Modeling: EngergyPlus at Ames Research Center", May 2011, Brief by Lawrence Berkeley National Laboratory.
5) "Ozone Widget Development Standards", Publication of NuWave Solutions.
6) Hauser, A., Verstege, J., "Seeing Global Power System States with Compact Visualization Techniques", date and publication reference unknown.

1.5. Overview of Document

This software requirements specification document provides a section which provides the overall description of the functionality of the product. It describes the informal requirements and is used to establish a context for the technical requirements specifications that are defined later in the document.

The Requirements Specification section of this document is written primarily for the developers and describes in technical terms the details of the functionality of the product.

2.0. Overall Description

2.1 System Environment

Overall use case for the EFFIS-INT Dashboard system is shown in Figure 1 below. The dashboard serves as the visualization component to the EFFIS-INT system. The EFFIS-INT system taps into energy data which in part is analyzed and performs optimizations on energy usage. The energy data along with optimization information is then sent to the EFFIS-INT dashboard for monitoring (situational awareness), conceptualization of alternatives (new courses of action), and end user control of the energy systems if deemed appropriate.

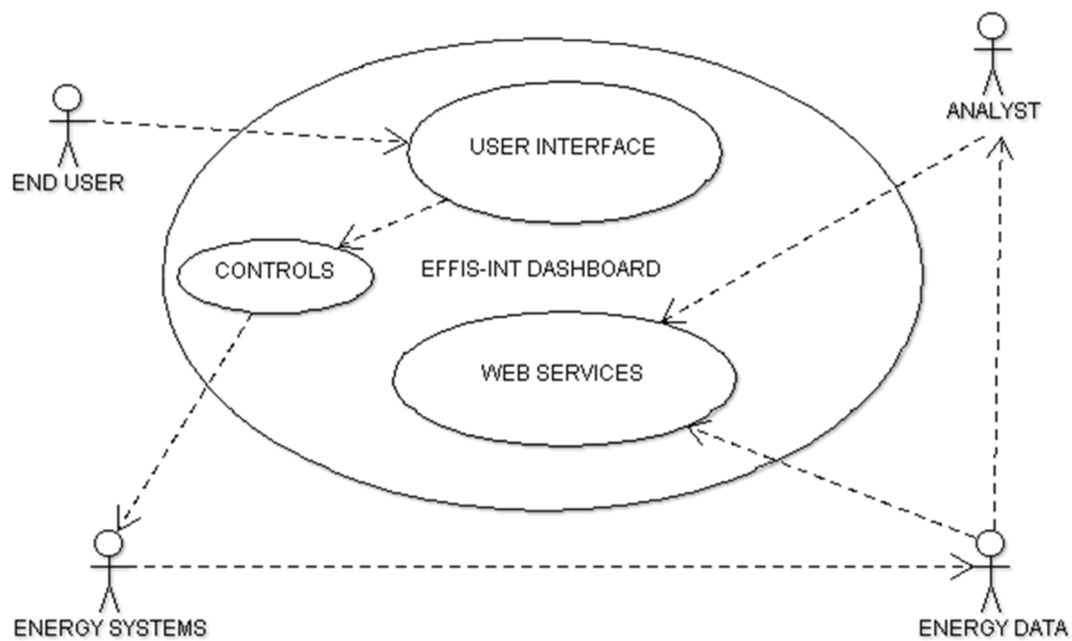


Figure 1 - System Environment

The EFFIS-INT Dashboard has in general one or more active actors, one or more data source systems along with one or more analysis systems. The active actors or end users utilize the system to monitor energy based information of C2 systems via the dashboard. The data source systems provide the capabilities to tap into and capture energy related data from the C2 system components. The analysis systems take in the energy based data and provide the capability to process the data and determine if there exists any relationships, patterns or indications of importance in the data in terms of means to optimize the energy usage, computational systems meeting operational requirements, and potential cyber threats.

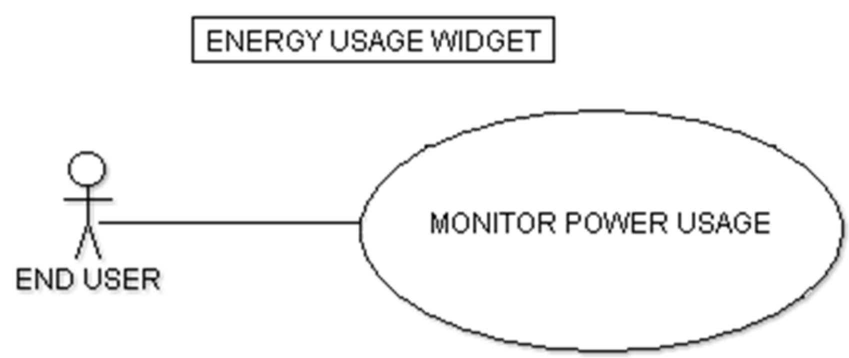
2.2 Functional Requirements Specification

The various use cases for the end users of the EFFIS-INT dashboard are described in separate sections. Each use case provides functionality available in the EFFIS-INT dashboard.

2.2.1 Monitor Power Consumption Use Case

Use case: **Monitor Energy Usage of System Power Components**

Diagram:



Brief Description
The user reviews and monitors power component widgets that show past, current, and projected power consumption.

Initial Step-By-Step Description
Before this use case can be initiated, the user has integrated the power component consumption widget(s) into the OWF framework for the given OWF portal and associated layout.

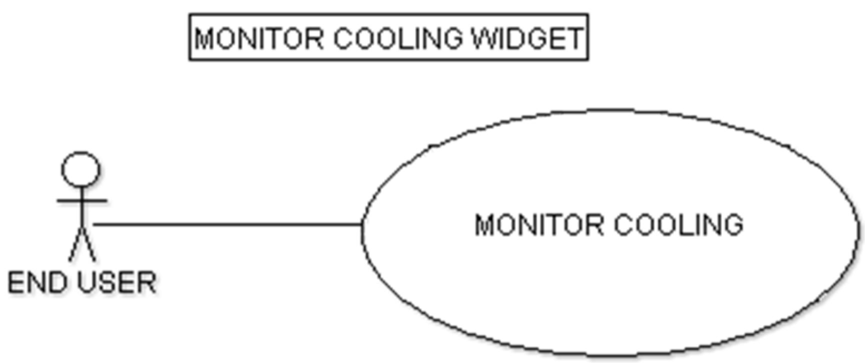
1. The user chooses to review power consumption of system components.
2. The system displays the past, current, and projected future energy consumption rates in the associated widget.
3. The user utilizes the display to understand the temporal variance of power consumption relative to current setting of desired power consumption.
4. The system presents alternatives to the current course of action based upon analytical processing.
5. The user assesses any necessary changes in power consumption to enter in control widgets.
6. The system provides continuous update to power component widgets.

Xref: 3.2.1

2.2.2 Monitor Cooling Use Case

Use case: **Monitor Cooling of System Components**

Diagram:



Brief Description
The user reviews and monitors cooling of system component widgets that show past, current, and projected cooling levels.

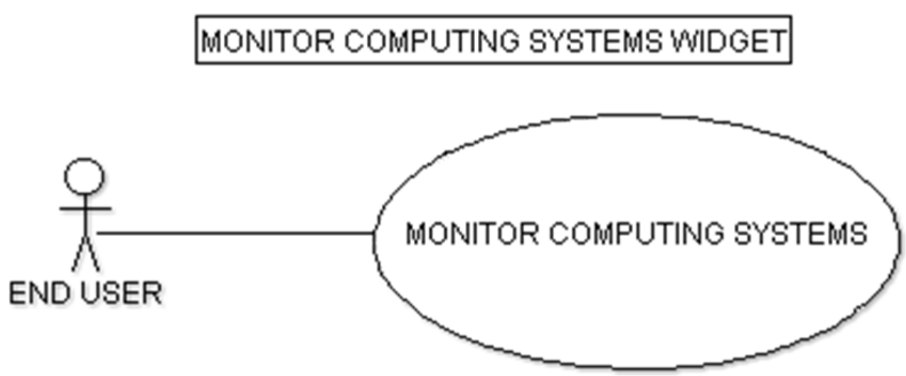
Initial Step-By-Step Description
Before this use case can be initiated, the user has integrated the cooling usage of power component widget(s) into the OWF framework for the given OWF portal and associated layout.

1. The user chooses to review the cooling of system components.
2. The system displays the past, current, and projected future cooling levels in the associated widget.
3. The user utilizes the display to understand the temporal variance of cooling levels relative to current setting of desired power consumption.
4. The system presents alternatives to the current course of action based upon analytical processing.
5. The user assesses any necessary changes in cooling levels to enter in control widgets.
6. The system provides continuous update to cooling widgets.

Xref: 3.2.2

2.2.3 Monitor Computing Systems Use Case
Use case: **Monitor Computing Systems**

Diagram:



Brief Description
The user reviews and monitors the computing systems widgets that show past, current, and projected energy usage of computing systems.

Initial Step-By-Step Description

Before this use case can be initiated, the user has integrated the computing systems energy usage widget(s) into the OWF framework for the given OWF portal and associated layout.

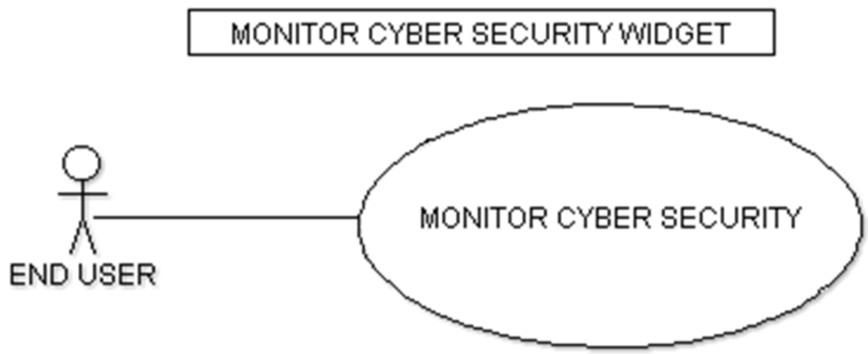
1. The user chooses to review the widget of computing systems.
2. The system displays the past, current, and projected future computing systems energy use in the associated widget.
3. The user utilizes the display to understand the temporal variance of computing systems power usage relative to current setting of desired power consumption.
4. The system presents alternatives to the current course of action based upon analytical processing.
5. The user assesses any necessary changes in power distribution to computing system and enters changes in the control widgets.
6. The system provides continuous updates in power provided to computing systems.

Xref: 3.2.3

2.2.4 Monitor Cyber Security Use Case

Use case: **Monitor Cyber Security**

Diagram:



Brief Description
The user reviews and monitors the cyber security widgets that show past, current, and projected energy usage conditions that might indicate cyber threats.

Initial Step-By-Step Description
Before this use case can be initiated, the user has integrated the cyber security widget(s) into the OWF framework for the given OWF portal and associated layout.

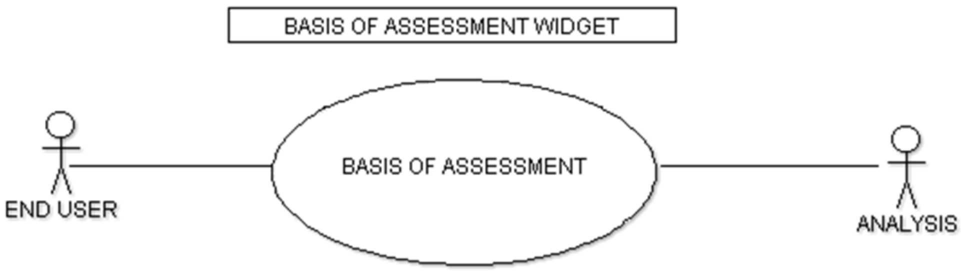
1. The user chooses to review the widget of cyber security.
2. The system displays the past and current cyber threat status or computations in the associated widget.
3. The user utilizes the display to understand the potential or actual cyber threats.
4. The system presents alternatives to the current course of action based upon analytical processing.
5. The user assesses any necessary precautions to combat actual or suspected cyber threats.
6. The system provides continuous updates in monitoring of potential cyber threats.

Xref: 3.2.4

2.2.5 Basis of Assessment Use Case

Use case: **Basis of Assessment**

Diagram:



Brief Description
Based on the selection of item in an EFFIS-INT dashboard monitor widget, the basis of assessment widget displays information relative to the selected item in a monitor widget. This relative information is the machine processed reasons regarding the assessment relative to the selected item. The user reviews the basis of assessment widgets that show reasons for the conditions shown in the monitor widget selected.

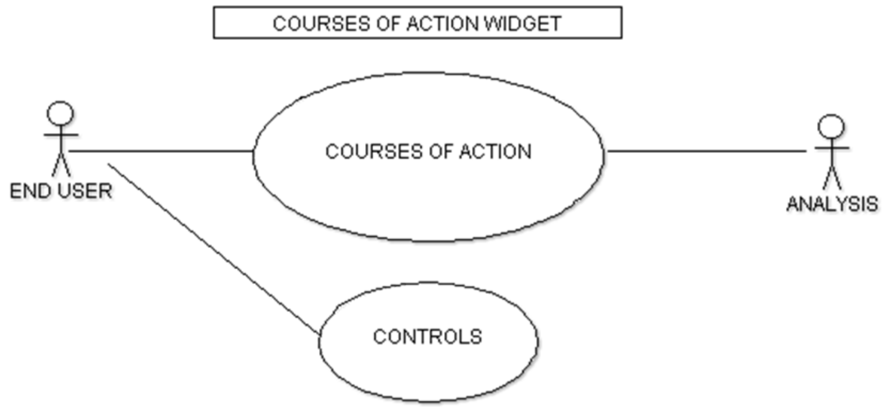
Initial Step-By-Step Description
Before this use case can be initiated, the user has integrated the basis of assessment widget(s) into the OWF framework for the given OWF portal and associated layout.

1. The user chooses to review the widget of basis of assessment.
2. The system displays the reasons or conditions that were applied to the state of the energy based data item selected in one of the monitoring widgets.
3. The user utilizes the display to understand the reasons and conditions associated with the machine assessment of the given energy data item.
4. The system provides the processed or policy based conditions, rules, or states in a visual form including rollover popup drill down regarding the assessments.
5. The user can quick review the same type of basis of assessments for any other energy data item found in any other monitoring based widget.
6. The system provides continuous updates to the assessments to maintain the current status of the assessment given any potential change to the associated energy data item.

Xref: 3.2.5
2.2.6 Courses of Action Use Case

Use case: **Courses of Action**

Diagram:



Brief Description

Based on the selection of item in an EFFIS-INT dashboard monitor widget, the basis of assessment widget displays information relative to the selected item in a monitor widget. This relative information is the machine processed reasons regarding the assessment relative to the selected item. The user reviews the basis of assessment widgets that show reasons for the conditions shown in the monitor widget selected.

Initial Step-By-Step Description

Before this use case can be initiated, the user has integrated the basis of assessment widget(s) into the OWF framework for the given OWF portal and associated layout.

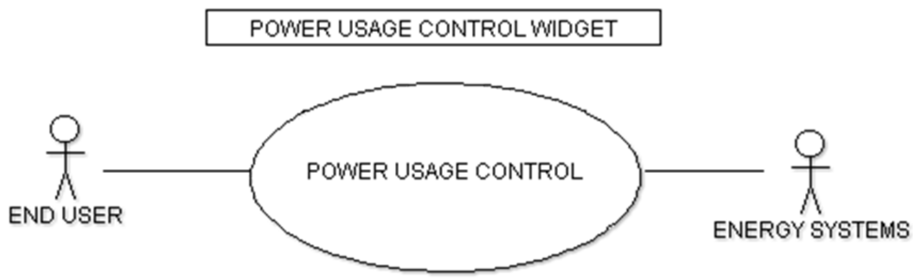
1. The user chooses to review the widget of basis of assessment.
2. The system displays the reasons or conditions that were applied to the state of the energy based data item selected in one of the monitoring widgets.
3. The user utilizes the display to understand the reasons and conditions associated with the machine assessment of the given energy data item.
4. The system provides the processed or policy based conditions, rules, or states in a visual form including rollover popup drill down regarding the assessments.
5. The user can quick review the same type of basis of assessments for any other energy data item found in any other monitoring based widget.
6. The system provides continuous updates to the assessments to maintain the current status of the assessment given any potential change to the associated energy data item.

Xref: 3.2.6

2.2.7 Power Usage Control Use Case

Use case: **Power Usage Control**

Diagram:



Brief Description

The user via the power usage control widget is able to adjust power usage settings either by entering their own levels of usage or via the selection of courses of action alternatives. The user selects the power component to which the control levels are set.

Initial Step-By-Step Description

Before this use case can be initiated, the user has integrated the power usage control widget(s) into the OWF framework for the given OWF portal and associated layout.

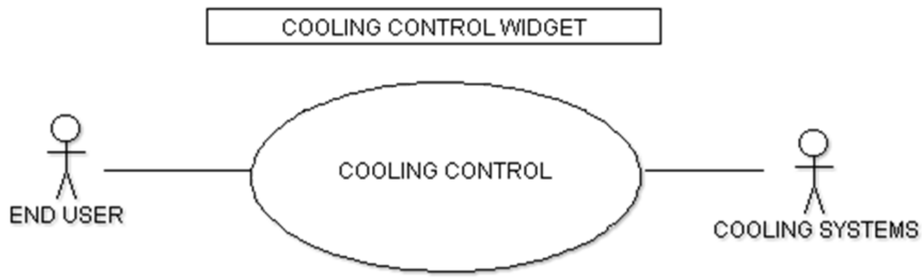
1. For the user to set the power usage levels and opens or views the power usage control widget.
2. The user selects one power component at a time and adjusts levels in the widget to modify its power usage settings.
3. The user may then repeat step 2 with another power component by selecting that component and performing the same adjustments.
4. This action provides feedback to the power usage monitor widget and the courses of action widgets to provide to update to the new settings.
5. The user is allowed to reset to existing level or keep the modifications.
6. The system updates the control unit to show the new selected levels of power usage in order to provide confirmation of the users actions.

Xref: 3.2.7

2.2.8 Cooling Control Use Case

Use case: **Cooling Control**

Diagram:



Brief Description
The user via the cooling control widget is able to adjust cooling settings either by entering their own levels of cooling or via the selection of courses of action alternatives. The user selects the cooling component to which the control levels are set.

Initial Step-By-Step Description
Before this use case can be initiated, the user has integrated the cooling control widget(s) into the OWF framework for the given OWF portal and associated layout.

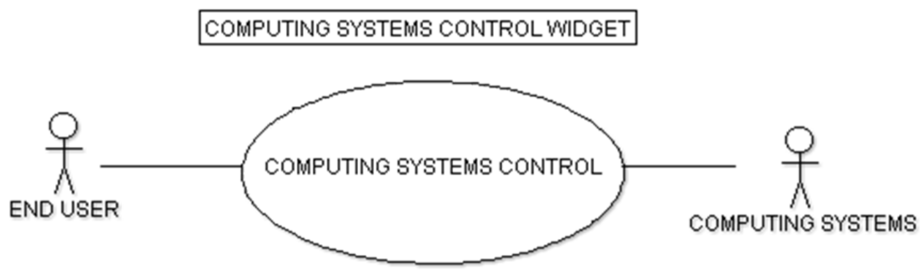
1. For the user to set the cooling levels, the opens or views the cooling control widget.
2. The user selects one cooling component at a time and adjusts levels in the widget to modify its cooling settings.
3. The user may then repeat step 2 with another cooling component by selecting that component and performing the same adjustments.
4. This action provides feedback to the cooling monitor widget and the courses of action widgets to provide to update to the new settings.
5. The user is allowed to reset to existing level or keep the modifications.
6. The system updates the control unit to show the new selected levels of cooling in order to provide confirmation of the users actions.

Xref: 3.2.8

2.2.9 Computing Systems Control Use Case

Use case: **Computing Systems Control**

Diagram:



Brief Description
The user via the computing systems control widget is able to adjust the various computing system settings either by entering their own levels of computing system adjustments or via the selection of courses of action alternatives. The user selects the computing system component to which the control levels are set.

Initial Step-By-Step Description
Before this use case can be initiated, the user has integrated the computing systems control widget(s) into the OWF framework for the given OWF portal and associated layout.

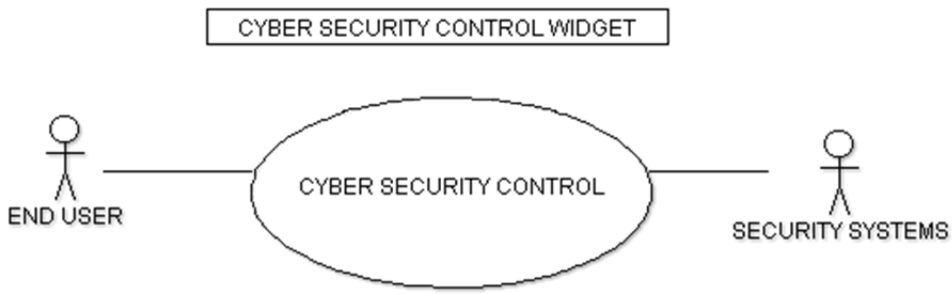
1. For the user to set the computing system levels, the user opens or views the computing system control widget.
2. The user selects one computing system component at a time and adjusts levels in the widget to modify its settings.
3. The user may then repeat step 2 with another computing system component by selecting that component and performing the same adjustments.
4. This action provides feedback to the computing system monitor widget and the courses of action widgets to provide to update to the new settings.
5. The user is allowed to reset to existing level or keep the modifications.
6. The system updates the control unit to show the new selected levels of the computing system in order to provide confirmation of the users actions.

Xref: 3.2.9

2.2.10 Cyber Security Control Use Case

Use case: **Cyber Security Control**

Diagram:



Brief Description
The user via the cyber security control widget is able to adjust the various cyber security settings either by entering their own levels of cyber security adjustments or via the selection of courses of action alternatives. The user selects the cyber security component to which the control levels are set.

Initial Step-By-Step Description
Before this use case can be initiated, the user has integrated the cyber security control widget(s) into the OWF framework for the given OWF portal and associated layout.

1. For the user to set the cyber security levels, the user opens or views the cyber security control widget.
2. The user selects one cyber security component at a time and adjusts levels in the widget to modify its settings.
3. The user may then repeat step 2 with another cyber security component by selecting that component and performing the same adjustments.
4. This action provides feedback to the cyber security monitor widget and the courses of action widgets to provide to update to the new settings.
5. The user is allowed to reset to existing level or keep the modifications.
6. The system updates the control unit to show the new selected levels of the cyber security component in order to provide confirmation of the user's actions.

Xref: 3.2.10

2.3 User Characteristics

The user is expected to be an experienced user of the C2 system along with knowledge and experience of the power usage requirements associated with the C2 system and its operational and policy requirements.

The user is expected to be allowed to make adjustments to power, cooling, computing, and cyber security settings via the EFFIS-INT dashboard widgets. The user is also expected to be allowed to accept or reject system recommendations via the courses of action recommendations.

The user may be different individuals assigned to perform the stated tasks involved with the EFFIS-INT dashboard, however for reasons of simplicity it is assumed that each user will have the same privileges associated with the control and setting of power, cooling, computing, and cyber security settings.

2.4 Non-Functional Requirements

The EFFIS-INT dashboard will be on a server with high speed Internet connectivity which is directly tied in to associated EFFIS-INT energy data source components as well as the associated C2 system(s) of interest. The physical machine will be required to conform to the same configurations as imposed upon the C2 system since it is tightly couple to information from this system. The software developed requires the use of a tool such as Tomcat and connection between supported components.

Web service interfaces shall exist between the various EFFIS-INT components and the C2 system(s) in order to allow the flow of energy related data to the analysis and visualization (dashboard) components. This same service also provides feedback of user actions via the control widgets.

3.0. Requirements Specification

3.1 External Interface Requirements

The EFFIS-INT dashboard widgets shall be connected via web services to the other EFFIS-INT components being the power, cyber, computing and cooling data collection system as well as the analysis and processing components of that energy data. This web service communication shall provide communication capabilities to and from each of the EFFIS-INT components. The assumption is made that the other EFFIS-INT components have been allows access to and have the proper interface to obtain the energy data from the C2 system of interest. A further assumption is made that the EFFIS-INT dashboard control widgets will be allowed to send energy level control settings to the C2 system of interest and that system will be able to respond appropriately as well as provide feedback to the results of the interaction.

The EFFIS-INT components shall also be connected via a two way web services communication with the C2 system(s) for which the EFFIS-INT system is performing energy optimization work. Control of the energy components shall be permitted as well via these web service connections. The assumption here is that the means of communication is via web services and the C2 system is capable of providing publish and subscribe transactions with the EFFIS-INT system. The EFFIS-INT system data exchange will also occur via hardware devices which collect the energy data. The assumption is that this data can wrapped into the web services data exchange.

3.2 Functional Requirements

3.2.1 Monitor Power Consumption

Use Case Name	Monitor Power Consumption
XRef	Section 2.2.1, Monitor Power Consumption
Trigger	The user assesses the monitor power widget
Precondition	The monitor power widget is installed in OWF of the dashboard and filled with energy usage data from the C2 system
Basic Path	1. The user opens or views the Monitor Power widget using the OWF methods as integrated into the dashboard. 2. The user reviews the various graphs and displays of the widget with each one representative of a particular component in the C2 system that provides energy data. 3. The user can select, rollover or hover over a given graphic element of the display to be presented with a popup or tooltip of amplifying information. 4. The widget is updated per an acceptable update rate for both the C2 and EFFIS-INT systems. 5. The widget provides the current baseline graphic for acceptable or set level of energy usage . 6. The widget in future version shall provide the capability to view alternative courses of action (COA) to the current set level of energy use. In addition, historical data relative to the energy usage shall be available.
Alternative Paths	There is currently only one path for a user to use the energy usage widget.
Postcondition	The user has a current understanding of the power usage by the various components of the C2 system. The user is also aware of how these values compare to the stated goals of energy usage.
Exception Paths	The courses of action or basis of assessment widgets will also provide situational awareness related to energy data of the C2 system. This, however is tightly coupled to the Monitor Power Consumption widget.
Other	Assumption of access to C2 energy data

3.2.2 Monitor Cooling

Use Case Name	Monitor Cooling
XRef	Section 2.2.1, Monitor Cooling
Trigger	The user assesses the monitor cooling widget
Precondition	The monitor cooling widget is installed in OWF of the dashboard and filled with energy usage data from the C2 system
Basic Path	1. The user opens or views the Monitor Cooling widget using the OWF methods as integrated into the dashboard. 2. The user reviews the various graphs and displays of the widget with each one representative of a particular component in the C2 system that provides cooling data. 3. The user can select, rollover or hover over a given graphic element of the display to be presented with a popup or tooltip of amplifying information. 4. The widget is updated per an acceptable update rate for both the C2 and EFFIS-INT systems. 5. The widget provides the current baseline graphic for acceptable or set level of energy usage . 6. The widget in future version shall provide the capability to view alternative courses of action (COA) to the current set level of cooling. In addition, historical data relative to cooling usage shall be available.
Alternative Paths	There is currently only one path for a user to use the energy usage widget.
Postcondition	The user has a current understanding of cooling by the various components of the C2 system. The user is also aware of how these values compare to the stated goals of cooling.
Exception Paths	The courses of action or basis of assessment widgets will also provide situational awareness related to cooling data of the C2 system. This, however is tightly coupled to the Monitor Cooling widget.
Other	Assumption of access to C2 energy data

3.2.3 Monitor Computing Systems

Use Case Name	Monitor Computing Systems
XRef	Section 2.2.3, Monitor Computing Systems
Trigger	The user assesses the monitor computing systems widget
Precondition	The monitor cooling widget is installed in OWF of the dashboard and filled with energy usage data from the C2 system
Basic Path	1. The user opens or views the Monitor Computing Systems widget using the OWF methods as integrated into the dashboard. 2. The user reviews the various graphs and displays of the widget with each one representative of a particular computing system in the C2 system that provides related data. 3. The user can select, rollover or hover over a given graphic element of the display to be presented with a popup or tooltip of amplifying information. 4. The widget is updated per an acceptable update rate for both the C2 and EFFIS-INT systems. 5. The widget provides the current baseline graphic for acceptable or set level of energy usage for the computing systems. 6. The widget in future version shall provide the capability to view alternative courses of action (COA) to the current set level of cooling. In addition, historical energy data relative to the computing systems shall be available.
Alternative Paths	There is currently only one path for a user to use the energy usage of computing systems widget.
Postcondition	The user has a current understanding of computing system energy needs by the various components of the C2 system. The user is also aware of how these values compare to the stated goals of the energy needed by the computing systems.
Exception Paths	The courses of action or basis of assessment widgets will also provide situational awareness related to computing systems data of the C2 system. This, however is tightly coupled to the Monitor Computing Systems widget.
Other	Assumption of access to C2 energy data

3.2.4 Monitor Cyber Security

Use Case Name	Monitor Cyber Security
XRef	Section 2.2.4, Monitor Cyber Security
Trigger	The user assesses the monitor cyber security widget
Precondition	The monitor cyber security widget is installed in OWF of the dashboard and filled with energy usage data from the C2 system
Basic Path	1. The user opens or views the Monitor Cyber Security widget using the OWF methods as integrated into the dashboard. 2. The user reviews the various graphs and displays of the widget with each one representative of a particular cyber security component in the C2 system that provides related data. 3. The user can select, rollover or hover over a given graphic element of the display to be presented with a popup or tooltip of amplifying information. 4. The widget is updated per an acceptable update rate for both the C2 and EFFIS-INT systems. 5. The widget provides the current baseline graphic for acceptable or set level of cyber security for the computing systems. 6. The widget in future versions shall provide the capability to view alternative courses of action (COA) to the current set level of cyber security. In addition, historical energy data relative to cyber security shall be available.
Alternative Paths	There is currently only one path for a user to use the montitor cyber security widget.
Postcondition	The user has a current understanding of computing system security needs by the various components of the C2 system. The user is also aware of how these values compare to the stated goals of the cyber security needed by the computing systems.
Exception Paths	The courses of action or basis of assessment widgets will also provide situational awareness related to energy data of the C2 system. This, however is tightly coupled to the Monitor Cyber Security widget.
Other	Assumption of access to C2 energy data

3.2.5 Basis of Assessment

Use Case Name	Basis of Assessment
XRef	Section 2.2.5, Basis of Assessment
Trigger	The user assesses the basis of assessment widget
Precondition	The basis of assessment widget is installed in OWF of the dashboard and filled with energy usage data from the C2 system
Basic Path	1. The user opens or views the Basis of Assessment widget using the OWF methods as integrated into the dashboard. 2. The user reviews the various graphs and displays of the widget with each one representative of a reasons for he current conditions of energy usage in the C2 system that provides related data. 3. The user can select, rollover or hover over a given graphic element of the display to be presented with a popup or tooltip of amplifying information. 4. The widget is updated per an acceptable update rate for both the C2 and EFFIS-INT systems. 5. The widget indicates the current baseline level of energy usage as part of the comparison in the assessment. 6. The widget in future version shall provide the capability to view alternative courses of action (COA) basis of assessment. In addition, historical basis of assessment reasons shall be available.
Alternative Paths	There is currently only one path for a user to use the basis of assessment widget.
Postcondition	The user has a current understanding of why the computing system energy usage is at the current point. The user is also aware of how these values compare to the stated goals of the energy needed by the computing systems.
Exception Paths	No exception paths because the basis of assessment is provided for only the reasoning behind the other widgets that are displayed in the dashboard.
Other	Assumption of access to C2 energy data

3.2.6 Courses of Action

Use Case Name	Courses of Action
XRef	Section 2.2.6, Courses of Action
Trigger	The user assesses the courses of action widget
Precondition	The courses of action widget is installed in OWF of the dashboard and filled with energy usage data from the C2 system
Basic Path	1. The user opens or views the Courses of Action widget using the OWF methods as integrated into the dashboard. 2. The user reviews the various presentations of alternative courses of action in the widget with each one representative of a analyzed and computed course of action regarding energy data of the system. 3. The user can select, rollover or hover over a given graphic element of the display to be presented with a popup or tooltip of amplifying information. 4. The widget is updated per an acceptable update rate for both the C2 and EFFIS-INT systems. 5. The widget provides the current baseline course of action for acceptable or set level of energy usage for the computing systems. 6. In addition basis of assessments will be provided for each course of action recommendation.
Alternative Paths	There is currently only one path for a user to use the courese of action widget.
Postcondition	The user has a current and potential new paths to take with regards to understanding of computing system energy needs by the various components of the C2 system. The user is also aware of how the presented alternative courses of action compare to the stated goals of the energy needed by the computing systems.
Exception Paths	No exception paths for courses of action other than to stay with the current course of action.
Other	Assumption of access to C2 energy data

3.2.7 Power Usage Control

Use Case Name	Power Usage Control
XRef	Section 2.2.7, Power Usage Control
Trigger	The user assesses the power usage control widget
Precondition	The power usage control widget is installed in OWF of the dashboard and filled with energy usage data from the C2 system
Basic Path	1. The user opens or views the Power Usage Control widget using the OWF methods as integrated into the dashboard. 2. The user reviews the various graphs and displays of the widget with each one representative of a particular control of the power to the C2 system. 3. The user can select, rollover or hover over a given graphic element of the display to be presented with a popup or tooltip of amplifying information. 4. The widget is updated per an acceptable update rate for both the C2 and EFFIS-INT systems. 5. The widget provides the current baseline graphic for acceptable or set level of energy usage for the computing systems. 6. The widget allows the user to set power usage levels for the various components and receive verification of the new settings.
Alternative Paths	There is currently only one path for a user to use the power usage control widget other than controlling power via other provided means in the C2 system.
Postcondition	The user has a current understanding of computing system energy needs by the various components of the C2 system and has been able to modify control settings via this widget based upon current conditions or upon recommendations from the courses of action widget.
Exception Paths	No exception paths other than using normal controls outside of the EFFIS-INT dashboard.
Other	Assumption of access to C2 energy data

3.2.8 Cooling Control

Use Case Name	Cooling Control
XRef	Section 2.2.8, Cooling Control
Trigger	The user assesses the cooling control widget
Precondition	The cooling control widget is installed in OWF of the dashboard and filled with energy usage data from the C2 system
Basic Path	1. The user opens or views the Monitor Computing Systems widget using the OWF methods as integrated into the dashboard. 2. The user reviews the various graphs and displays of the widget with each one representative of a particular computing system in the C2 system that provides related data. 3. The user can select, rollover or hover over a given graphic element of the display to be presented with a popup or tooltip of amplifying information. 4. The widget is updated per an acceptable update rate for both the C2 and EFFIS-INT systems. 5. The widget provides the current baseline graphic for acceptable or set level of energy usage for the computing systems. 6. The widget allows the user to set power usage levels for the various components and receive verification of the new settings.
Alternative Paths	There is currently only one path for a user to use the energy usage of computing systems widget other than controlling power via other provided means in the C2 system.
Postcondition	The user has a current understanding of computing system energy needs by the various components of the C2 system and has been able to modify control settings via this widget based upon current conditions or upon recommendations from the courses of action widget.
Exception Paths	No exception paths other than using normal controls outside of the EFFIS-INT dashboard.
Other	Assumption of access to C2 energy data

3.2.9 Computing Systems Control

Use Case Name	Computing Systems Control
XRef	Section 2.2.9, Monitor Computing Systems Control
Trigger	The user assesses the computing systems control widget
Precondition	The computing systems control widget is installed in OWF of the dashboard and filled with energy usage data from the C2 system
Basic Path	1. The user opens or views the Monitor Computing Systems widget using the OWF methods as integrated into the dashboard. 2. The user reviews the various graphs and displays of the widget with each one representative of a particular computing system in the C2 system that provides related data. 3. The user can select, rollover or hover over a given graphic element of the display to be presented with a popup or tooltip of amplifying information. 4. The widget is updated per an acceptable update rate for both the C2 and EFFIS-INT systems. 5. The widget provides the current baseline graphic for acceptable or set level of energy usage for the computing systems. 6. The widget allows the user to set power usage levels for the various components and receive verification of the new settings.
Alternative Paths	There is currently only one path for a user to use the energy usage of computing systems widget other than controlling power via other provided means in the C2 system.
Postcondition	The user has a current understanding of computing system energy needs by the various components of the C2 system and has been able to modify control settings via this widget based upon current conditions or upon recommendations from the courses of action widget.
Exception Paths	No exception paths other than using normal controls outside of the EFFIS-INT dashboard.
Other	Assumption of access to C2 energy data

3.2.10 Cyber Security Control

Use Case Name	Cyber Security Control
XRef	Section 2.2.10, Monitor Cyber Security Systems
Trigger	The user assesses the cyber security control widget
Precondition	The cyber security control widget is installed in OWF of the dashboard and filled with energy usage data from the C2 system
Basic Path	1. The user opens or views the Monitor Computing Systems widget using the OWF methods as integrated into the dashboard. 2. The user reviews the various graphs and displays of the widget with each one representative of a particular computing system in the C2 system that provides related data. 3. The user can select, rollover or hover over a given graphic element of the display to be presented with a popup or tooltip of amplifying information. 4. The widget is updated per an acceptable update rate for both the C2 and EFFIS-INT systems. 5. The widget provides the current baseline graphic for acceptable or set level of energy usage for the computing systems. 6. The widget allows the user to set power usage levels for the various components and receive verification of the new settings.
Alternative Paths	There is currently only one path for a user to use the energy usage of computing systems widget other than controlling power via other provided means in the C2 system.
Postcondition	The user has a current understanding of computing system energy needs by the various components of the C2 system and has been able to modify control settings via this widget based upon current conditions or upon recommendations from the courses of action widget.
Exception Paths	No exception paths other than using normal controls outside of the EFFIS-INT dashboard.
Other	Assumption of access to C2 energy data

3.3 Detailed Non-Functional Requirements

3.3.1 Logical Structure of the Data

The logical structure of the data to be sent and received via web services to the EFFIS-INT dashboard overall data flow or structure is outlined in the figure below.

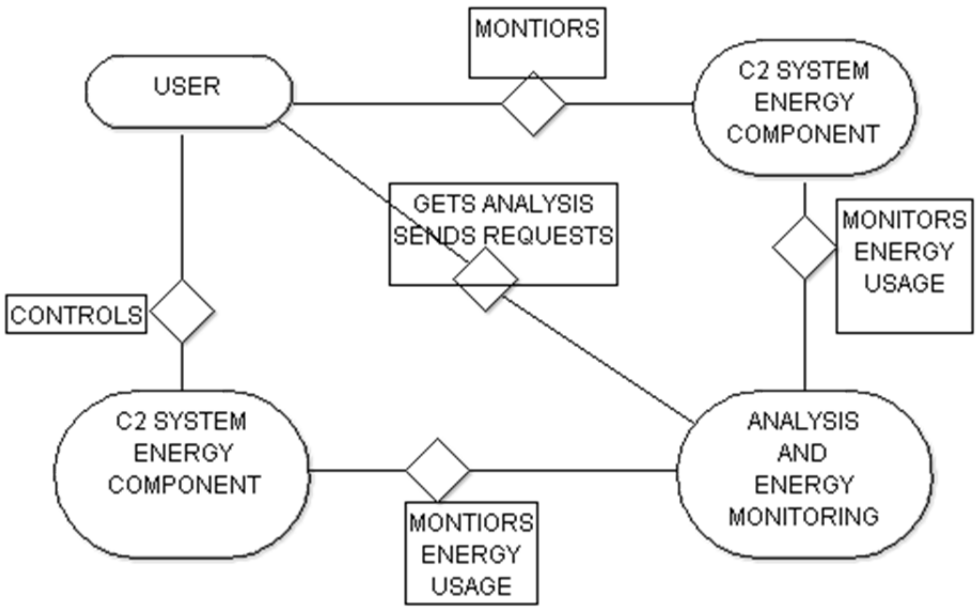


Figure 2 - Logical Structure of the EFFIS-INT Dashboard

The data descriptions of each of these data entities is still to be defined and is relative to the intended C2 system of use. These details would be descriptive of the C2 system and its associated energy components as well as energy monitoring devices utilized. For most cases the details employ an associated name and id for a given data item, its origin and destination, the action to take place, the user or system who performed the action, a date and time stamp, any comments or inputs related to the action, required time frame for action, security classification on message, list of components effected by or to which the action or message applies, the response time required, the alert level, previous conditions, current conditions, expected conditions, and general information of the action or message.

3.3.2 Security

The server on which the EFFIS-INT dashboard resides will have the security protection to match the C2 system to which it is assigned. In addition the OWF framework provides an additional layer of security which can be applied as an additional measure as well as a means to control access to only those users authorized to use the EFFIS-INT dashboard. This is critical when control permissions are given to this user, which is the user, will have the capability to control critical systems within the C2 system domain. Access for publish and subscribe to the web services of the data sources shall be limited to the EFFIS-INT system.

Hosting the EFFIS-INT system, which includes the dashboard within the same network allows for improved security control and access via the web services. Each C2 system that utilizes the EFFIS-INT system shall define it's specific security requirements which may impact features of the EFFIS-INT dashboard if the data sources are made unavailable by security restrictions in access.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-01-0188

1. REPORT DATE *(DD-MM-YYYY)*
September 2013

2. REPORT TYPE
Capability Investment: First Year

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE

Energy-Focused Fusion Information System Integration
A NISE funded Capability Investment Project
Year 1 FY 2013 Report

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHORS

Stacey Curtis
Adriane Wotawa-Bergen Wolfe

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

SSC Pacific, 53560 Hull Street, San Diego, CA 92152–5001

8. PERFORMING ORGANIZATION REPORT NUMBER

TD 3278

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Innovative Science and Engineering Program
SSC Pacific, 53560 Hull Street, San Diego, CA 92152–5001

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The objective of our project is to develop sustainable science and technology in command, control communications, computers, and intelligence (C4I) energy and energy management. Sustainability development speaks to development that can be sustained without compromising the future. The three pillars of sustainable development are economics, equity, and environment. In this project, the three pillars of a capability investment were conceptualized to map to the pillars of a sustainable business development strategy.

The specific goals are to conduct technical development of a command and control (C2) energy management system for Navy C4I/information technology (IT) systems. Three tasks were identified to achieve this development: (1) investigate computing, cooling, energy management and cyber–physical security, (2) optimize the system to reduce energy usage while maintaining IT performance and security with the target of a 50% reduction in total system energy consumption, and (3) develop a C2 energy management approach for cross-platform Navy end-users. We have two business development goals: (1) grow a sustainable S&T business area related to this project, and (2) transition the EFFIS-INT project approach to C4I-IT systems. The goal of workforce development is to develop people who are technically knowledgeable in the field so that SSC Pacific can become a center of excellence in IT-energy management.

15. SUBJECT TERMS

workforce development	business development	efficiency management	end-user security
techn ology development	data center systems	machine learing	

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	U	107	Stacey Curtis
					19B. TELEPHONE NUMBER *(Include area code)* (619) 553-5255

INITIAL DISTRIBUTION

84300	Library	(2)
85300	Archive/Stock	(1)
71760	Stacey Curtis	(1)

Defense Technical Information Center
Fort Belvoir, VA 22060–6218 (1)

SSC Pacific
San Diego, CA 92152-5001